



Investigation of Quantum-inspired Modelling in Interactive
Search based on Information Foraging Theory

Dr Amit Kumar Jaiswal

This is a digitised version of a dissertation submitted to the
University of Bedfordshire.

It is available to view only.

This item is subject to copyright.

Investigation of Quantum-inspired Modelling in Interactive Search based on Information Foraging Theory



Amit Kumar Jaiswal

Supervisor(s): Haiming Liu, Ingo Frommholz

Examining Committee: Stefan Rieger, Gareth J.F. Jones

PhD Candidate Student Number: 1808389

Registration Date: 02/04/2018

Current Degree and Mode of Study: PhD, Full Time

Research Institute: IRAC

This dissertation is submitted for the degree of
Doctor of Philosophy

Dedicated to my parents, Anita and Late Chandrashekhar

Abstract

This thesis investigates the use of the mathematical formalism of quantum mechanics for modelling users' information needs from the viewpoint of Information Foraging Theory (IFT). IFT has been successfully applied to model user behaviours and preferences in the information retrieval (IR) process, which motivates this work that hypothesises IFT can enhance user interaction and the effectiveness of typical IR and recommendation tasks, such as multimodal query auto-completion and image recommendation. During interactive IR sessions, users' information needs often evolve, which requires more system assistance and support to capture these dynamics. The users' information needs in such an interactive session compared to the typical unambiguous query terms tend to be multi-semantic and heuristic in the way of natural languages. In an effort to solve this problem, an interactive multimedia and multimodal IR system is developed based on a quantum-inspired mathematical framework utilising Hilbert spaces. Based on IFT, the key methodology involves characterising the users' multimodal information needs. The users' multimodal information needs are integrated into the IR system through a projective transformation that follows mathematical constructs of the quantum probabilistic framework. The proposed quantum-inspired interactive framework is evaluated through an image retrieval task, which allows a multi-iteration of textual queries for finding specific images in a search session supported by auto query completion and visual query cues. Our main findings are: in an interactive multimodal IR context, multi-semantic queries effectively help to confirm users' information needs in a session; from the spatial context of IR, our framework captures the dynamics of evolving information needs and reflects the historical interaction of the multimodal IR process. This dissertation, therefore, provides comprehensive insights and findings about the usage of the quantum probabilistic framework in interactive IR. It exhibits the usefulness of the quantum-inspired IR framework to fine-grained user aspects based on IFT, enhancing the representation and modelling constructs to explicit the information retrieval behaviours and preferences.

Acknowledgements

I am grateful to my supervisor Haiming Liu, who has been truly inspirational throughout my candidature. She always shows me a tireless enthusiasm to meet, discuss, listen and encourage. Her invaluable pieces of advice and faith make all the difference, and I have been blessed to find in her all the good qualities of a supervisor. Many thanks must go to my co-supervisor Ingo Frommholz, who has given me the opportunity to work in this field of research and extra guidance, constant encouragement, and constructive feedback that substantiate to be invaluable in improving the quality of my work.

I would also be thankful to the faculty at the University of Bedfordshire's Institute for Research in Applicable Computing has provided me with a nice working place with all much-needed facilities and services, as well as hosting seminars and the Postgraduate research forum for considerable research highlights. I am also grateful to RGS, especially Caroline Aird, who has provided me assistance in this research endeavour.

My sincere appreciation goes to Stefan Rieger and Gareth J.F. Jones for their time and expertise to examine this thesis.

I am indebted to the European Union's Horizon 2020 research and innovation programme for funding my research generously, under the Marie Skłodowska-Curie Innovative Training Network fellowship. A special thanks to the QUARTZ project consortium for training and research support, and numerous project meetups across Europe and Britain. I am also thankful to my colleagues from QUARTZ Young Radicals Group, particularly Sagar Uprety, Prayag Tiwari, and Benyou Wang, to share and discuss generic research, ideas, and feelings, and above all their friendship.

I have been fortunate to have an excellent mentor during my secondment visit, Massimo Melucci at the Department of Information Engineering, University of Padova. I enjoyed the discussions with Massimo and his group members, Emanuele Di Buccio, Benyou Wang, and Qiuchi Li, to which I am thankful.

Last but most importantly, the constant support and encouragement of my family, friends, and colleagues had been inevitable throughout my Ph.D. Especially, I am highly obliged to my mother Anita Jaiswal, and forever grateful for the encouragement that you gave me, and

for supporting me unconditionally and spiritually throughout this candidature and my life in general.

Origins

The following publications co-authored by me have derived as part of this dissertation:

- **Chapter 3**

- **Amit Kumar Jaiswal**, Guilherme Holdack, Ingo Frommholz, & Haiming Liu [95], “Quantum-like Generalization of Complex Word Embedding: a lightweight approach for textual classification”, in *Proceedings of the Conference "Lernen, Wissen, Daten, Analysen" (LWDA)*, 2018
- **Amit Kumar Jaiswal** [94], “Investigating Interactive Information Retrieval via Information Foraging Theory”, in *Proceedings of FDIA Workshop, ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR)*, 2018
- **Amit Kumar Jaiswal**, Haiming Liu, Ingo Frommholz [98], “Reinforcement Learning-driven Information Seeking: A Quantum Probabilistic Approach”, in *Proceedings of Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS) Workshop, SIGIR*, 2020

- **Chapter 4**

- **Amit Kumar Jaiswal**, Haiming Liu, & Ingo Frommholz [96], “Effects of foraging in personalized content-based image recommendation”, in *Proceedings of The 2nd International Workshop on Explainable Recommendation and Search (EARS), SIGIR*, 2019
- **Amit Kumar Jaiswal**, Haiming Liu, & Ingo Frommholz [97], “Information Foraging for Enhancing Implicit Feedback in Content-based Image Recommendation”, in *Proceedings of the 11th Forum for Information Retrieval Evaluation (FIRE)*, pages 65-69, 2019

- **Chapter 5**

- **Amit Kumar Jaiswal**, Haiming Liu, & Ingo Frommholz [99], “Utilising Information Foraging Theory for User Interaction with Image Query Auto-Completion”, in *Pro-*

ceedings of the 42nd European Conference on Information Retrieval, pages 666-680, Springer, Cham, 2020

- **Chapter 6**

- **Amit Kumar Jaiswal**, Haiming Liu, & Ingo Frommholz [100], "Semantic Hilbert Space for Interactive Image Retrieval", in *Proceedings of the ACM SIGIR International Conference on the Theory of Information Retrieval*, 2021.

Declaration

I, Amit Kumar Jaiswal, declare that this thesis is my own work except where explicitly referenced otherwise in the text and that the contents of this dissertation are original and have not been submitted for consideration of any other degree or professional qualification except as specified. This dissertation is my own work and contains the outcome of work done with, or with the assistance of others, which is specified as such.

Amit Kumar Jaiswal
June, 2023

Table of contents

Origins	ix
List of figures	xvii
List of tables	xix
Nomenclature	xxi
1 Introduction	1
1.1 Motivation for Using Quantum-inspired IR Approaches and Information Foraging Theory	4
1.2 Research Questions	7
1.3 Contributions	10
1.4 Dissertation Overview	11
1.5 Notations	13
2 Background and Literature Review	15
2.1 Prologue to Information Retrieval	15
2.1.1 Information Retrieval Models	16
2.1.2 Interactive Search	18
2.1.3 Reinforcement Learning for IR	19
2.1.4 Query Auto-Completion	21
2.2 Introduction to Information Foraging Theory	21
2.2.1 Key Constructs	22
2.2.2 Role of IFT in IR	23
2.2.3 Language Model for Behavioural IR	27
2.3 Desiderata of Quantum-inspired IR Models	29
2.3.1 Preliminaries	30
2.3.2 Introduction to Quantum Probability	41

2.3.3	Probabilistic Nature in Quantum IR Models	45
2.3.4	Generalisation of Language Embeddings	46
2.3.5	Generalised Classical Language Models	47
2.3.6	User Interactions	47
2.3.7	Realm of IFT	48
2.4	Conclusion	50
3	Using Quantum-inspired Word Embedding for Modelling Searcher Behaviour	51
3.1	Generalising Word Embedding using Quantum Probabilistic Framework . . .	51
3.1.1	Related Work	52
3.1.2	Method	53
3.1.3	Experiments	57
3.1.4	Results	58
3.1.5	Conclusion	60
3.2	Quantum-inspired Reinforcement Learning-driven Information Seeker . . .	60
3.2.1	Related Work	62
3.2.2	Information Seeker as Reinforcement Learning Agent - Hypothesis	63
3.2.3	The Framework	64
3.2.4	Conclusion	74
4	Modelling Searcher Preferences in Content-based Image Recommendation based on Information Foraging Theory	77
4.1	Overview	77
4.2	Related Work	79
4.3	Personalised Content-based Image Recommendation	81
4.4	Experiment I - Content-based Image Recommendation using Information Foraging-driven Interventions	83
4.4.1	Users' Attention based on Visual Information Foraging	84
4.4.2	Data	84
4.4.3	Results	85
4.5	Experiment II - Implicit Feedback in Content-based Image Recommendation using Information Foraging Theory	86
4.5.1	Implicit Feedback based on Information Scent	86
4.5.2	Data	87
4.5.3	Training	88
4.5.4	Results	89
4.6	Conclusion	91

5	User Information Needs in Image Query Auto-Completion based on Information Foraging Theory, using Language Model	93
5.1	Overview	93
5.2	Related Work	95
5.2.1	Query Auto-Completion	95
5.2.2	Query Suggestion	96
5.2.3	Information Foraging Theory	96
5.2.4	Language Embeddings	97
5.3	Problem Formulation	97
5.4	Model	98
5.4.1	Image Query Auto-Completion	98
5.4.2	iBERT for Patch Probability	100
5.5	Information Foraging Perspective on Visual Information Needs	101
5.5.1	Patch Selection	101
5.6	Experiments	104
5.6.1	Datasets	104
5.6.2	Training	104
5.6.3	Performance Measure	105
5.7	Results	105
5.8	Conclusion	107
6	Quantum-inspired Modelling for Interactive Image Retrieval	109
6.1	Semantic Hilbert Space for Interactive Image Retrieval	109
6.1.1	Related Work	113
6.1.2	The Framework - Semantic Hilbert Space	115
6.1.3	Model	118
6.1.4	Experiments	123
6.1.5	Results	125
6.1.6	Conclusion	128
6.2	Quantum Probabilistic Modelling for Query Interaction in Image Search . .	128
6.2.1	Quantum-inspired Interactive Model	130
6.2.2	Experiments	133
6.2.3	Results	134
6.2.4	Conclusion	135

7	Conclusions	137
7.1	Summary of Work	137
7.2	Main Findings and Limitations	139
7.3	Future Work	142
7.3.1	User Interaction in Image Query Auto-Completion	143
7.3.2	Incorporating User Behavioural Features in IR System	143
	References	145
	Appendix A Ethical Approval	165
A.1	Application Form for Pilot User Study	165
A.2	Participant Consent Form	170
	Appendix B Datasets	173
B.1	Access Link	173

List of figures

1.1	Dissertation Overview	12
2.1	IR process in a Reinforcement Learning Framework based on the general scenario	20
2.2	Overview of Information Patch	23
2.3	Overview of Information Scent	24
2.4	Hilbert Space	44
3.1	Selection of Documents in Hilbert Space	66
3.2	Architecture of Quantum-inspired reinforcement learning framework	67
3.3	Pictorial representation of a single action and it's corresponding unitary transformation for quantum state update in a Hilbert space	74
4.1	Schematic Architecture of Personalised Recommender System	81
4.2	User Interface of Content-based Image Recommender System	82
4.3	WikiArt Image Dataset	88
4.4	Prediction Performance	90
5.1	An instance of Image Query auto-completion using our extended LSTM language model	95
5.2	The end-to-end architecture of Image Query Auto-Completion.	99
5.3	IFT based Explanation of Image Query Auto-Completion	102
6.1	A pictorial representation of a word 'date' which shows multiple meanings depending on the grounded context	110
6.2	An instance of our formulated retrieval task which depicts the expressive nature of user information needs in the context of an image retrieval scenario.	111
6.3	An overview of the proposed Semantic Hilbert space framework.	115
6.4	Complex convolution layer based on Trabelsi et al. [206]	121
6.5	An instance of sample training set images from the MIT States dataset.	123

6.6	An instance of sample training set images from the Fashion200k dataset. . .	124
6.7	Some qualitative examples of our proposed model. The examples are from MIT States dataset.	127
6.8	Some qualitative examples of our proposed model. The examples are from Fashion200k dataset.	127
6.9	Example retrieval from the test set based on a multimodal query against a plain textual query	127
6.10	An end-to-end architecture of Quantum-inspired Interactive model.	129

List of tables

1.1	Relationship between IR and IFT	7
1.2	List of Notations	14
3.1	List of the pre-trained word embedding models where †depicts GloVe embeddings and ‡depicts Fasttext embeddings. The word embeddings vector is of 300 dimension.	58
3.2	Overview of the Datasets	58
3.3	Evaluation Results across different word embedding models. ‘R’ depicts the updated word embedding model with word representation as real-valued numbers. And, ‘M’ depicts the reproduced embedding mixture based on [124]	59
3.4	Training time per epoch for each word-embedding model across varied text classification datasets listed in Table 3.2	59
4.1	Recommendation Result in terms of information scent scores	85
4.2	Classification Report	90
5.1	Evaluation results of the query completion task. Our MRR score is in bold face.	106
5.2	Perplexity of image query auto-completion on both datasets utilising an image and indiscriminate noise. Inclusion of image results in a better (lower) perplexity	106
6.1	Performance comparison with baselines. (+) depicts the BERT model added to the TIRG for a new baseline. The best performances are in boldface . . .	125
6.2	Ablation test on both MIT States and Fashion200k datasets. (+) and (-) depicts with and without the respective components to the proposed SHS mode.	126
6.3	Specifications of the Image and Query Representations	134
6.4	Performance on Visual Genome where query representations is of same (*) and varied memory size(**)	134

Nomenclature

Acronyms / Abbreviations

ACT-R Adaptive Control of Thought-Rational

AOS Accuracy of Scent

AUROC Area under the Receiver Operating Characteristic

BERT Bidirectional Encoder Representations from Transformers

CNN Convolutional Neural Network

CTVA CODE Theory of Visual Attention

DNN Deep Neural Networks

GloVe Global Vectors of Word Representation

HRED Hierarchical Recurrent Encoder-Decoder

iBERT Image Bidirectional Encoder Representations from Transformers

IFT Information Foraging Theory

IR Information Retrieval

ISE Information goal, Search strategy, Evaluation threshold

IS Information Scent

ISL Information Scent Level

ISP Information Scent Pattern

IUNIS Inferring User Need by Information Scent

LSTM	Long Short-Term Memory
MDP	Markov Decision Process
MPC	Most Popular Completion
NLP	Natural Language Processing
OM	Ostensive Model
PCA	Principal Component Analysis
PCBIR	Personalised Content-based Image Recommendation
PMI	Pointwise Mutual Information
PPSM	Probabilistic Patch Selection Model
PT	Projective Transformation
QAC	Query Auto-Completion
QIIM	Quantum-inspired Interactive Model
QLM	Quantum Language Model
QM	Quantum Mechanics
qRL	Quantum-inspired Reinforcement Learning
QT	Quantum Theory
QWeb	Quantum World Wide Web
ReLU	Rectified Linear Unit
RL	Reinforcement Learning
RNN	Recurrent Neural Network.
RPN	Region Proposal Network
SERP	Search Engine Result Page
SHS	Semantic Hilbert Space
SVM	Scalable Vector Machine

TF-IDF Term Frequency - Inverse Document Frequency

UI User interface

URN Unidirectional Recurrent Network

Chapter 1

Introduction

Today, search engines are a key driver for billions of people for finding information across the globe. There are numerous forms of information content such as texts, images, videos, audio, and multimodal documents [19] scattered across the web, and accessible to users in a manner able to satisfy their information needs [22]. The multitude of users' information needs is in stiff competition with search engines. However, searchers explicate their information needs differently such as via a textual or visual query (images), voice sample, and video, and so the notion of information retrieval (IR) arises to address users' information needs [53].

Understanding searchers' continually changing information needs [231, 72] during the use of web search engines is one of the most challenging tasks in information retrieval. At present, web search engines have transformed how people search for information. The way people seek an answer to certain information via a keyword-based query on the Web enriches historical interaction between search engines and their users. When search engines are not able to interpret a user's query, they provide query suggestions [37] or web links (or sponsored search options such as advertisement) for predicting result usefulness and redirect the user to the search engine result page (SERP). Such rich interaction will potentially help users to reach episodic information faster. Interpreting what the user is looking for and where (university name, university address, geographical coordinates, etc.), even if it be external (e.g. location, time of the day) or internal (e.g. user interests) helps find the context that additionally defines the search as well as critical to rank and display resources will be an added advantage to future search systems. The issue of modelling information needs that elicits users' changing aspects (such as the cognitive state), specifically since IR models appear to have attained accomplishment and there is an explicit need to go beyond the current state-of-the-art. Nowadays, the entire web search stack enriches historical user interaction / real-time data, emerging from crawling strategies to renovating the presentation of the results [103, 107, 126, 155, 224]. We note that the users are generally reluctant to provide

substantial information to depict an unambiguous information need [154, 147, 211]. Also, users do not always perceive how to express their information needs, and they rarely have only an indistinct knowledge of what they are seeking. On the other hand, users' knowledge and interests might alter during the search process, by tweaking their information needs. Apart from the standard relevance feedback models, such as the Okapi model [178] or the Rocchio algorithm [179], few recent works have tried to record context [140] or interaction [194]. Moreover, there is an important need for a principled framework that integrates both, and then, equally important attempts to capture the different manifests of possible interaction or user behaviour, namely, search intention, query reformulation, clicks, and navigation. Those tasks are all effectuated frequently in web searches.

IR has been extensively applied to model users' information needs through different approaches such as [235, 66, 36, 42, 24, 130, 187] in varied tasks. However, these traditional IR models lack a perceptual or generalised framework that can encapsulate evolving information needs. Thus, Ingwersen and Järvelin [91] introduced a geometrical framework that delineates multiple facets of user behaviour (information need, feedback, etc.) and capable of incorporating them. This led to a set of certain models [171, 170, 63] developed to model users' information needs. These approaches belong to formal models of information retrieval and act as an inbuilt framework that elicits user behaviours [171, 63, 170, 122, 209] such as interaction, feedback, etc. These formal models of IR are based on the mathematical formalism of quantum mechanics i.e., Hilbert space formalism which uses quantum probability, and hence the modelling process becomes instinctively complex whilst incorporating or undertaking user behaviour. Additionally, in classical IR, a theoretical framework of dynamic IR [197] was introduced that delineates user interaction based on the incorporated probabilistic models [66]. However, this framework focuses on the users' information goal after following their interaction which limits the generalising ability (such as the input query to be an image instead of texts) as compared to quantum-inspired IR models. Inspired by these frameworks, we employ behavioural framework of IR - Information Foraging Theory [166] to incorporate user aspects such as their behaviours in quantum-inspired models.

We note that users' information needs in the recommender system are user profiles that characterise their preferences. Recommender systems can be personalised which requires predetermination of users' information needs. In general, the user information needs are textual queries, however, visual content such as an image can be referred to as visual information need [73, 167]. Information Foraging Theory is an exclusive framework for enhancing users' ability to search and so an IR system can be improved. Thus, finding the users' preferences via implicit feedback can improve image recommendation. The context of users' information needs can be distinct such as in query suggestions, where the search

engine presents a list of suggestions that the user follows. This scenario can be translated to a more granular task of query auto-completion, where a text prefix competes with the search engine to generate a complete query. However, it can be slightly extended to bring the context of distinct information needs for image search, where auto-completed query renders user attention and this can be useful to enhance interactive IR. This also reflects the contextual diversity [172, 158] of user's information need.

In this thesis, we focus on characterising and modelling users' information needs with the usage of quantum-inspired IR models from the viewpoint of Information Foraging Theory, where a searcher is treated as a forager. A certain set of information retrieval tasks have been considered to cater to the aforementioned problem. Initially, a formulation is developed for tackling cognitive aspects between words based on the quantum probabilistic framework that formalises the classical embedding components that drive users' information needs. This quantum-inspired formulation is incorporated within a reinforcement learning model to describe the strategy on how a forager's actions can be learned in a query matching task. This led to a theoretical quantum-inspired reinforcement learning framework that can guide an information seeker (or forager) who has no prior information about the search environment. Based on the constructs of IFT, we then tackle to designate user preferences in an image recommender system, where the system attempts to attract searchers toward recommended items in order to capture their information needs. The method uses IFT to delineate users' preferences and also considered it differently as a form of implicit feedback that enhances the overall capability of the recommendation system. To further this work, a query auto-completion task has been undertaken for image search, where we cater to users' distinct information needs. Specifically, the query auto-completion uses a character-based language model for predicting the query and we employ it as an input to the iBERT model for estimating the probabilities of image patches. These image patches entirely form a single image to render the user's attention (or contextual diversity of the distinct information need). Consequently, this thesis presents two quantum-inspired frameworks for (interactive) image retrieval tasks, where in one of the frameworks the user's information needs is interactive i.e., a textual-visual query, and the second framework encapsulates a set of distinct information needs as an input. We find that the interactive representation of users' information needs can renovate the image retrieval task implemented in a quantum-inspired framework.

1.1 Motivation for Using Quantum-inspired IR Approaches and Information Foraging Theory

The role of users' in an information retrieval system is regarded as a major player in driving the search engine. Due to the necessity of keeping users' in a loop, we anticipate a better retrieval system to manifest certain behavioural aspects of a searcher. We have considered certain aspects of a user and their challenges highlighted in the previous section. The formulation of any quantum theory-based methods/frameworks for IR systems should be signified by these inclined aspects. The notion of applied quantum theory-based models (or quantum probabilistic models) [170, 199] is employed to enhance interactive IR systems, especially the evolving and vague user's information needs based on Information Foraging Theory. The behavioural aspects of a searcher are designated using IFT. My work is to renovate the user-oriented IR systems, especially the user's information needs by generalising the classical approaches of it with the usage of quantum theory. Realistically, the genesis of quantum theory (QT) is in Physics and it is employed as a mathematical formalism that follows quantum probability to model the behaviours and dynamics of microscopic particles. To be precise, the quantum theory here refers to 'quantum information theory' and the underlying mathematical framework is borrowed from 'quantum mechanics' (QM), where the notion of its' usage in IR is mathematical instead of physical [211]. The key reason for using such a mathematical framework that can depict the QM resemblance in IR, for instance, a user searches 'nearby Asian restaurant' and follow the list of retrieved results, however, the user may specify their initial information need or confer feedback to the search system by declaring restaurants that are relevant. This process of IR is interactive and so the user behaviour alters the state of the system (or the system itself). This entirely leads to the instinctive role of users perform in the search process and rooted information has been demonstrated to be effectively acquired mathematically. And, the information retrieval community refers to it as quantum-inspired IR [210]. The underlying mathematical framework of quantum mechanics has distinct properties of their components such as state vector/eigenvector, observable, superposition, and compatibility forms identical notions in IR. The representation of information objects (texts, images, etc.) uses the Hilbert space formalism, to delineate the vector space models in classical IR by a complex-valued vector space i.e. Hilbert space [211].

In this thesis, the work deals with leveraging the mathematical framework of quantum mechanics to model the user's information needs based on IFT. Also, the introduced quantum-inspired framework in this thesis neither pertains to quantum machine learning, nor is it pertained to quantum physics. Rather it uses certain approaches developed in quantum-

inspired IR [211]. Considering the aforementioned points, I inscribe the motivation behind the usage of the QM-based mathematical framework (or Hilbert space formalism) to model users' semantic information needs to be based on IFT. It acts as an instinctive indicator as to why QM and IFT are key contenders for integrating the user's behavioural aspects in an IR system.

Evolution of Information Needs are based on Information Foraging Theory

Users' information needs are explicated by a textual or visual query depending upon the type of search system (text search or image search). However, searchers having under-specified information need [154] encounter it difficult to describe textually what it is they are seeking. In the case of image search, the search engine reclines to be more exploratory as compared to the text search. The text-based queries are either short or ambiguous or broad and lack cognitive aspects of user [147], whereas image queries as IN manifest the user beliefs [228] and tend to be even shorter [73]. Image search depends vastly on user interaction with SERPs (retrieved images) which renders user intention (or user beliefs). From the viewpoint of Information Foraging Theory, a forager possess certain action which is described as information need (assume it is a textual query) and follows the outcomes (search results) in a patch-wise fashion. Each of the retrieved results is unique information patches characterising the goal (information diet or information goal) of a forager, and these patches are traced via cues provided there must exist an information scent. The information scent changes rapidly at each point of time during foraging (or seeking) [237] and it is due to the fact that going through information patches, there exist numerous cues around and forager (or searcher) choose one among it, and so it augments the current state of the information scent that searcher possesses. Alternatively, if considered an image search, then the retrieved images may consist of multiple objects (or sub-patches) within an image which delineate a pattern (or distribution) of information scent [157].

Generally, in IR, the user's information needs are unclear initially and then after certain modifications, it reaches the information goal. However, we considered some edge cases of information needs in the previous paragraph, where not only the user's query alters but also search results make an impact on the user's decision, and that let their information need to be modified. This reflects a perpetual tendency of the searcher to be incognito for search results that are unexplored and relinquished. The advantage of using the IFT constructs, especially information scent can be employed to characterise the manifold of information needs in tasks, such as query auto-completion and recommender systems.

Variability in User Information Needs

In IR, the underlying user's information needs in a search task enlist multiple outcomes (search results). However, after a certain refinement of the retrieved results, the user selects two potential results that have equal chances of being the information goal and there realises the situation of uncertainty. The uncertainty can be elaborated based on the existing work [237], where selecting one of the outcomes from two can involve risk or ambiguity and vice versa for the second outcome. The explanation of this uncertainty is folded with the IFT constructs, where risk refers to the merit of the outcome (prevalent information patch) and the user's segmented perception of it. This type of uncertainty indicates whether a specific information patch the user selects to exploit or set on, as this acts as a prior information patch. To minimise the risk behaviour, the searcher will permit exploitation over exploration, this leads the user to remain at a specific information patch and potentially leave out the rest of the distal patches. The second type of uncertainty is ambiguity which permits to not forage (or search) any other information patches. The other information patches refer to an unknown distribution due to unvisited search outcomes. This type of uncertainty usually occurs with users' information needs, however, I can recall from the previous motivation where information needs are evolving and vague to be modelled. This brings the perspective of representing such complex information needs using Hilbert space formalism, where an information need is uncertain and evolving/vague.

The general aim of our research is inspired by the prior work [171] established in [211], where they considered that the user information need become more specific while reading an abstract (or summaries), typing keywords, or clicking on a document with respect to a system or a user point of view. [119] suggested that a search system behaves differently if the content it contains is altered i.e., adding or removing documents, and user click behaviour updates accordingly (user starts clicking on new documents or stops clicking on the removed ones). Following the notion of the aforementioned work, Information Foraging has been applied in [156, 157] to understand web search behaviour by introducing the concept of information scent level (ISL) and information scent pattern (ISP). However, they did not consider the major factor of user behaviour which is query reformulation or information need as it is always influenced by varying a query.

Following the above elaboration of information needs, it can be stated that the user information need is not evident. It can be envisioned that several information needs are in a superposition state, in a situation where the user interacts with information patches (documents, images, etc.), and lead the information need to be specific. However, foraging

Table 1.1 Relationship between IR and IFT

IR / Information Seeking	Information Foraging
Searcher/Seeker	Forager
Image	Information Patch
Object within Image (or Perception Component)	Sub-patch or Forager's information diet
Identified object / SERP	Information diet
Image selection	Patch choice
User attention to object	Information scent level or Information diet choice
Disengaging from object/patch	Patch leave
Diverted attention	Weak information scent

behaviour follows a sequential decision making process and this can benefit the overall quantum-inspired framework if incorporated within the information need representation.

The listed instances of user information needs can be inherently modelled using the Hilbert space formalism. As mentioned above, foraging is a sequential decision making process and so is its influence toward bringing the user-centric facet. An analogy between IR and IFT is reported in Table 1.1. This reveals the behavioural aspect of IR and altogether forms an exciting task for quantum-inspired information retrieval as to whether it can be applied.

1.2 Research Questions

The major goal of this dissertation is to address the following broad research question:

RQ: *How can a mathematical language of quantum mechanics be applied in information retrieval, underlying the behavioural viewpoint of Information Foraging Theory, and in particular to the user information needs?*

Our approach to tackling this question is two-fold, (a) to investigate interactive phenomena in information retrieval tasks where behavioural aspect arises and can be enhanced using Information Foraging Theory, and (b) to integrate the underlying behavioural IR aspect that follows IFT in a quantum-inspired interactive framework.

To this aim, we employ the mathematical description of quantum mechanics to derive a model that exploits explicit (user preferences, query reformulation, or query completion) and implicit feedback, such as clicks, user behaviours (i.e. attention, navigation, etc.) based

on Information Foraging Theory. We begin with the consideration that existing studies of interactive IR frameworks [171, 170] assume that the user information needs to become more specific, whilst they read abstracts (or summaries), type keywords, or click on documents with respect to a system or from a user perspective. However, if the content of a search system changes (e.g. documents are added, altered, or removed), user behaviour updates accordingly (he/she starts clicking on new documents and avoids availing the removed ones) [119]. Also, various search algorithms (e.g., if documents are ranked differently) exhibit changes in user search behaviour [85]. The alteration in search results starts when the user clicks over it due to changing document relevance over time [56]. To enhance formal IR models [63, 90] for interactive search, we focus on a different form of information objects such as images other than text-based information needs due to the lack of cognitive aspects in keyword-based queries [147]. However, queries in image search manifest the user and tend to be even shorter [73].

Specifically, we investigate the answer to this question with the following sub-research questions:

RQ 1: How can a quantum probability view of word embedding be applied in information retrieval, and in particular to the information seeker?

The notion of word embeddings is to describe the semantic relationship among words. However, such embedding models find it difficult to capture the casual meaning of combined words such as a sentence. In the IR domain, the mathematical framework of quantum mechanics [211] has been employed for the representation of words and to capture the interaction among words [201, 241]. The first proposal to enhance classical word embedding is overlaid with a quantum probabilistic framework of IR [124]. We employ this framework to generalise the training of complex-valued word embedding with an algorithmic approach that optimises the model and is computationally effective. This generalised model is then employed to incorporate within an agent-based framework to guide information seekers (or foragers). To effectuate foragers (as searcher) [237] cognitive ability during the search, we employ Information Foraging Theory [166] to understand how searchers can learn in the ongoing process of finding information. Also, the learning ability of a searcher can be channelled by the reinforcement learning approach by giving a free choice of search scenarios in an uncertain environment. However, the search process is cost-driven, and assessing the incurred cost by a searcher within an uncertain environment can potentially optimise forager in finding the information. The conventional approach of modelling and analysing foraging behaviour is inspired by IFT which maximises the searcher's net amount of information

gained to the time allocated between and within information patches. User action and their action dynamics during search play an important role in changing behaviour and user beliefs state [228]. However, it has been recently demonstrated that action behaviour can be learned to represent using reinforcement learning [39] by extrapolating a policy into two components - action representation and its transformation. In Chapter 3, we elaborate on a generalised quantum-inspired word embedding model and present the behavioural reinforcement learning framework parameterised using constructs of quantum probability.

RQ 2: To what degree can user interaction mechanism be explained by Information Foraging Theory?

User interaction in Web search is a leading paradigm of context-based information retrieval systems. It has been extensively adopted to rectify general Web searches in query suggestion [221, 37], content-based image retrieval [129], result ranking [5, 243], query auto-completion [103, 126], etc. We, therefore, reckon that understanding rich user interaction in multimedia search scenarios will help determine the interactive elements to model the user information needs. Also, modelling user information needs requires an explainable and behavioural approach for the formal IR models to support which we endue through Information Foraging Theory. Here, we make an attempt to solicit through the lens of Information Foraging to understand user interaction in some of the general IR tasks. Chapter 4 and Chapter 5 elaborated certain IR tasks to answer this question.

RQ 3: Whether the user interaction mechanisms using Information Foraging Theory could inform effective formal (quantum) models for interactive search?

Interactive search possesses user interaction [90, 63, 129, 155, 17, 68] as an important part of the search process which enhances the search system and searchers, and also left an adverse effect on users' cognitive ability [30, 207]. Recent advances in Quantum theory, as a mathematical formalism to model the dynamics and interaction in quantum physics, have been adopted for incorporating user interactions as a comprehensive textual/visual representation [171, 63, 220, 201, 28, 223, 216, 252, 209], which leveraged the fundamental concepts of quantum theory: superposition and entanglement to delineate interactions. Inspired by the empirical gains of quantum theory, we adopt the quantum-inspired IR-based models as a generalisation of classical approaches to formulate rich interaction processes provided vague and evolving information needs based on IFT. In Chapter 6, I adapt the Hilbert space formalism to model the user multimodal information need (textual-visual query) in an image

search scenario to see whether a searcher’s cognitive aspects renovates. The searcher’s cognitive aspect is signalled via a visual query and so combined with a text query able to manifest the inherent information need.

1.3 Contributions

This dissertation investigates quantum-inspired IR-based approaches based on Information Foraging Theory, and on how IFT shapes behavioural aspects of IR. In this section, the key contributions of this thesis are summarised as follows.

1. As recent new derivatives of word embedding models have been developed, we leverage the complex-valued word representation for text classification tailored for language-dependent problems. We present an algorithmic approach to the existing complex-valued embedding model [124]. The algorithmic approach optimises the model in a way that makes the training steps computational efficient as well as a lightweight embedding model. This quantum-like word embedding model is then employed as a sub-network in the Actor-Critic reinforcement learning (RL) framework, where the complex-valued word embedding model generates reward values classified among positive, negative, and partial matches. The reinforcement learning framework uses quantum probability constructs to represent user aspects (query, state, etc.). This quantum-inspired RL framework is developed to guide a forager during the information seeking process by means of *Reinforced Foraging*.
2. We propose an Information Foraging based strategy to investigate the explicit (user preferences) and implicit feedback (users’ attention) in an image recommender system. A content-based image recommendation system is presented that incorporates the user’s visual attention to recommended items. It also elicits how user preferences affect item selection on recommendation using information scent. Also, it magnifies implicit behavioural signals whilst user pick their interesting images ascertained via visual cues. The strategy provides information scent artifacts to evaluate the strength of visual cues for user-item interaction relevance.
3. We present a task of query auto-completion for image search, referred to as *Image query auto-completion*. The language model employed for predicting the next sequence of characters follows [93] and extends the functional user embeddings by image embeddings. Also, we propose an explainable approach for the user interaction behaviour based on Information Foraging Theory [166] to explicit the inspected

challenges of varying users' information needs. For the inspection of varied information needs, we present the iBERT (image Bidirectional Encoder Representations from Transformers) model inspired by [50] to reckon probabilities of patches.

4. We propose a quantum-inspired IR based model, refer to as *Semantic Hilbert space* for interactive image retrieval task. This framework models the dynamic information needs of the user, where an information need consists of the aligned textual-visual query. In this framework, we introduce a method that leverages the projective transformation in a complex-valued Hilbert space to delineate the encoding of still images via textual features. Another similar task is considered, where the user information needs are a set of queries for which, we present an interpretable framework inspired by quantum probabilistic framework for interactive image search which exploits patch-level captions as an aspect of weak supervision while training.

1.4 Dissertation Overview

I outline the structure of the dissertation shown in Figure 1.1.

In Chapter 3, I explore the quantum-inspired word embedding model which is used for text classification. I generalise the complex-valued word embedding model with an algorithmic approach that optimises the training steps. I utilise this embedding model in the Critic network of a theoretic quantum-inspired reinforcement learning framework for guiding the searcher in an information seeking process. This chapter derives answers for the research question **RQ 1**.

In Chapter 4, I investigate the notion of Information Foraging Theory in characterising searcher preferences in a content-based image recommender system. The IFT-based strategy elicits how user preferences affect item selection on the recommendation. It also improvises the preferences as implicit behavioural signals for user reinforces their attention while the selection of recommended items using information scent. Following this, we investigate how the information needs to shape the user's attention in a query auto-completion task applied in an image search. This is elaborated in Chapter 5. These two chapters derive answers for the research question **RQ 2**.

In Chapter 6, we follow the quantum probabilistic framework to address the problem of modelling user's information needs which are evolving and vague, in an image retrieval task. We develop a new engaging model using frameworks which reliant on Hilbert space formalism to represent multimodal IR. The expressiveness of the model lies in the fact that the user information needs are textual-visual queries, which are not fused but rather one of the feature spaces projects upon the other. The visual query in the input along with text

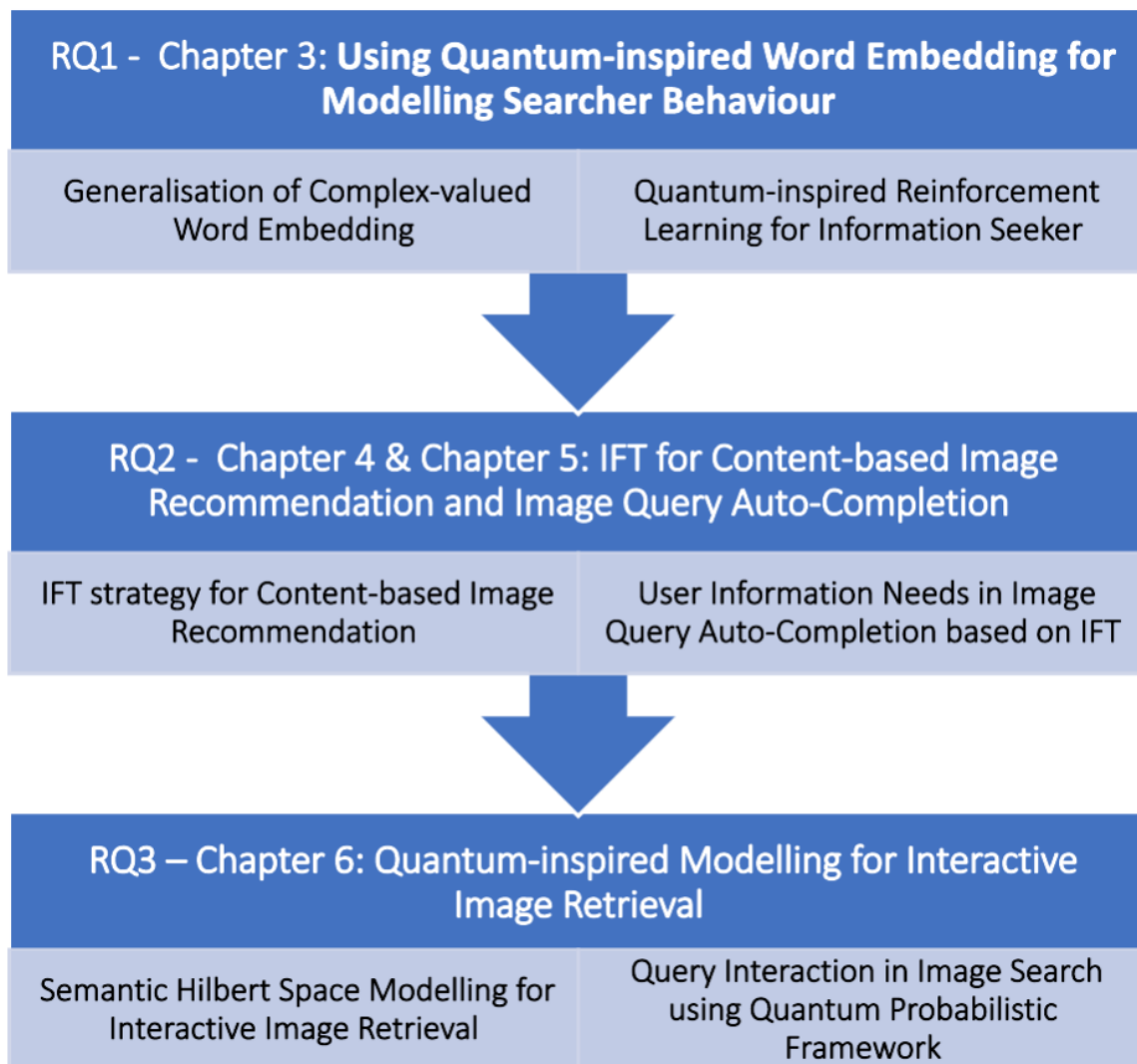


Fig. 1.1 Dissertation Overview

follows the IFT-based strategy developed for delineating the varying process of information needs in Chapter 5. We also present a projective transformation in the semantic Hilbert space model to characterise the encoding of still images through textual features. Having found that expressive information need are important for the robustness of image retrieval systems in terms of performance. This led to another similar task of image retrieval, where the user information needs are a set of queries in a session. The notion of varying information needs emanates from the dynamics of a user whilst altering the input query, which makes it a task of interactive IR. This chapter derives answers for the research question **RQ 3**.

Chapter 7 comprehensively answers the main finding of the aforementioned research questions, broader impacts on IR, and future directions following the work in this dissertation.

1.5 Notations

Throughout this entire thesis, we follow certain generic notations categorised based on the chapters. We follow a range of notations listed in Table 1.2. We employ lowercase γ to represent the dynamic factor for enumerating principal components. Some of the notations used here follows [46, 249]. The representation of orthonormal basis dimension of Hilbert space is denoted by b_i , the tensor product by \otimes , R denotes the rank of \mathcal{G} and \mathcal{L} and has n-order tensor of rank 1.

Table 1.2 List of Notations

Notation	Interpretation	Description
γ	Dynamic factor	Counter for identifying the irrelevant components
α_{i,b_i} $ \phi_{b_i}\rangle$	$b_i \in \{1, \dots, k\}$ Semantic meaning	Probability amplitude Basis vector (n (word vectors) or k^n dimension for tensor product of basis vectors)
$ w_i\rangle$ $ q_i\rangle$	$\sum_{b_i=1}^k \alpha_{i,b_i} \phi_{b_i}\rangle$ $ w_1\rangle \otimes w_2\rangle \dots \otimes w_n\rangle$	Word state vector Query state vector
$ \Psi_q^T\rangle$	$\sum_{b_1, \dots, b_n=1}^k \underbrace{\left(\prod_{i=1}^n \alpha_{i,b_i} \phi_{b_i}\rangle \otimes \dots \otimes \phi_{b_n}\rangle \right)}_{\mathcal{L}_{b_1, \dots, b_n}}$	Local representation (\mathcal{L} is a k^n dimensional tensor)
$ \Psi_q\rangle$	$\sum_{b_1, \dots, b_n=1}^k \mathcal{G}_{b_1 b_2 \dots b_n} \phi_{b_1}\rangle \otimes \dots \otimes \phi_{b_n}\rangle$	Global representation of combined meanings/patches
\mathcal{G}	$\sum_{r=1}^R w_r \cdot e_{r,1} \otimes e_{r,2} \otimes \dots \otimes e_{r,n}$	Probability amplitude (semantic space of meaning)
$\langle \Psi_q^T \Psi_q \rangle$	$\sum_{b_1, \dots, b_n=1}^k \underbrace{\mathcal{G}_{b_1 \dots b_n} \times \prod_{i=1}^n \alpha_{i,b_i}}_{\text{Probability amplitudes}}$	Projection of the global representation to the local representation of a query
$ a_i\rangle$	$(a_1\rangle, a_2\rangle, \dots, a_R\rangle)^T$	Output of the Actor network
\vec{Q}	$\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$	Set of textual queries
\vec{I}	$\{\vec{i}_1, \vec{i}_2, \dots, \vec{i}_n\}$	Set of visual (image) queries
\vec{R}	$\{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n\}$	Set of target image queries
$\mathcal{E}_p(\cdot)$	Visual features	Pre-trained image embedding model
$\mathcal{P}(q, i)$	$ q\rangle \langle i $	Projection of an input image query on a textual query as inner product among it
P_T	Image symmetry	Projective transformation
M		Mapping Function
M_I		Image Mapping Function
I_M	$Rot(P_T)M_I(i_f)$	Multimodal representation of input query
$g(i_f, q_f)$		Multimodal representation function
\mathbb{I}	$I + \iota Q$	Multimodal input matrix
\mathcal{L}_{CE}		Loss function of complex-valued encoder network
\mathcal{L}_{CD}		Loss function of complex-valued encoder network
ρ_q	$\sum_{i=1}^n q_i^2 f_i\rangle \langle f_i $	Query density

Chapter 2

Background and Literature Review

This chapter brings forward insight into the basics of information retrieval and methods, and a brief overview of certain IR tasks and their application. A highlight covering the notions of IR models, the mathematical framework of quantum mechanics, and user-oriented aspects of IR. To fulfil the need for user-oriented aspects in IR, we briefly explained Information Foraging Theory and certain IR tasks that involve user interaction, including IFT approaches in recommendation tasks. The literature review also spotlights the need for mathematical concepts that emerged in the quantum-inspired IR framework in general.

2.1 Prologue to Information Retrieval

In general, an information retrieval system is affined where the input content (documents, images, videos, etc.) is transformed into an appropriate representation by a certain indexing mechanism. In an IR system, a user input content is expressed via an information need that is shaped as a textual query. The process delineated by an IR system weighs the user query adverse available documents representation through a certain matching function based on the specific IR model. This step of comparison in an IR system generates a catalogue of documents presented to the user that are ranked accordingly, beginning with the most relevant to the least favoured. The IR system comprehends that the retrieved catalogue ratiocinates those documents are about the input query. The conviction of aboutness in IR tends to be semi-shaped based on the epitome character of queries and documents using terms [211]. This process of retrieval may adopt user feedback to differentiate documents' relevance based on their information need. The user information needs can be altered and so their behaviour changes the state of the IR system. This led to the birth of interactive IR.

In this section, we first briefly explain the concepts of classical IR models and then its prodigy - interactive information retrieval. Following this, we highlight the IR tasks that

involve the user aspect and finally, the mathematical framework of quantum mechanics that can describe the interactive IR process.

2.1.1 Information Retrieval Models

The establishment of IR models is to describe and formulate the representation of queries and documents, confer norms and algorithms to measure the documents' relevance, and therefore retrieve documents in decreasing order. IR models are prone to characterising information implied within queries and documents. The importance of these traditional IR models can be informed as crucial baselines to our new quantum-inspired IR-based approaches. We only describe IR models which are concerned and formed the basis of developed models in this dissertation.

Vector Space Model

This is the backbone of the similarity-based IR model, which is introduced to represent text documents in a way where the objects retrieved are modelled as elements of a vector space. The information objects include a term or sentence of a document or a query, where vectors represent the objects retrieved. The required constructs for this algebraic model are a set of documents and a set of query which can be framed as, $D_s = \{td_1, td_2, \dots, td_n\}$ and $Q_s = \{q_1, q_2, \dots, q_n\}$, where td and q depicts a textual document and a query. The elements of the corresponding query and document vectors are commonly allied with terms emerging in the test collection. The available terms within a test collection of a document are coordinates to the dimension of vector space. A weighting factor is used to formalise the values of the elements of every vector. There are several schemes of weighting, for instance, if one uses a binary weighting [48] then vector elements delineate a query or a document (with values zero or one). It represents the term presence or absence corresponding to the allied vector element. However, an alternative weighting method can be the term frequency - inverse document frequency (TF-IDF) [133, 109]. The TF-IDF elicits the terms' importance based on their occurrence in a document rather than in the test collection. The formula of the weighting constructs of vectors can be equated as:

$$td_i = TF_{i,td} \underbrace{\log_2 \frac{M}{N}}_{IDF} \quad (2.1)$$

where $TF_{i,td}$ depicts the term (t_i) frequency in document D_s , M depicts all documents in the collection and N represents those documents that endue term t_i . The log component in

Eq. 2.1 represents IDF. This weighting factor captures the terms' importance that resides in documents and queries. A list of weighting methods is comprehensively investigated in [184]. TF act as bag-of-words and IDF is a weighs by how often a word occurs in the corpus. It rates the rare words higher than the common words. The bag-of-words model is to elucidate text (i.e. a sentence or a document) which is represented as a dictionary of its words, disregarding grammar and even word order but keeping multiplicity (frequency).

The merit among documents D_s and queries Q_s can be measured as the similarity between corresponding vectors. It uses the inner product as

$$sim(D_s, Q_s) = \frac{\sum_{i=1}^n D_{s_i} \cdot Q_{s_i}}{\sqrt{\sum_{i=1}^n (D_{s_i})^2} \cdot \sqrt{\sum_{i=1}^n (Q_{s_i})^2}} \quad (2.2)$$

where the dimension of vector space is n under which the vectors are spanned. This $sim(.)$ represents the cosine of the angle θ . So, the cosine similarity in the vector space model supports a peculiar order of documents, in decreasing order of their similarity scores with a particular query.

Probabilistic Model The classical probabilistic IR model was initially introduced in [177] to identify the relevant documents based on the decreasing order of documents' relevance probability. The retrieval of relevant documents is considered a classification task [191]. Specifically, this retrieval task can be translated as a classification problem, where the set of documents $D_s \in \{r, \neg r\}$, r depicts the class of relevant documents and $\neg r$ depicts the class of irrelevant documents respectively. To estimate the relevance probability provided a document td , as $P(D_s = r|td)$ can be equated as

$$P(r|td) \equiv \frac{P(R|td)}{P(\neg r|td)} = \frac{P(td|r)P(r)}{P(td|\neg r)P(\neg r)} \equiv \frac{P(td|r)}{P(td|\neg r)} \quad (2.3)$$

Considering a scenario where the terms (t_i , where $i \in [1, \dots, n]$) implied in document td are conditionally independent. It follows the Eq. 2.3 to be refined as

$$\frac{P(td|r)}{P(td|\neg r)} \approx \prod_{i=1}^n \frac{P(t_i|r)}{P(t_i|\neg r)} \equiv \sum_{i=1}^n \log \frac{P(t_i|r)}{P(t_i|\neg r)} \quad (2.4)$$

Documents with their corresponding probabilities reflect that the user's action (or information need) elicits the likelihood of being a particular document relevant. Estimating the probabilities entail the terms that forms document among the classes of documents which are relevant and irrelevant. This kind of estimation broadens the horizon and a meaningful measure for IR tasks.

An extension to the probabilistic IR model [177, 66] is introduced for user-oriented information retrieval [67], in particular, to describe the process of user interaction and the evolving activities within it such as query suggestion/reformulation. We highlight the language model in Section 2.2 for the sake of conciseness with interactive IR.

2.1.2 Interactive Search

Evolving information needs are mostly intermittent to information retrieval systems as the user search criteria tend to be ambiguous and shape as time progresses. Such information needs require the IR system to be standardised, in particular, explicating ostensive retrieval. Ostensive information retrieval [36] is principally tailored for evolving and ambiguous information needs. The explication of Ostensive IR emerges with user-system interaction such as in distinct query formulation, where query shaping tends to assume prescribed interactions as more vital than others. Its ad-hoc nature stands as a theoretical basis for search engines without query functionality on the interface or for the method which delineates alteration in users' information needs. Alternatively, the user can subjugate by interpreting their information need to a short phrase. Ostensive IR offers a way to recuperate information need details unavoidably phased out whilst following an interaction technique that is considered instinctive with respect to the cognitive process.

Ostensive IR characterises the querying or feedback process by means of an 'intermediary action' among explicit feedback and implicit feedback. Ostensive IR follows an ostensive language that (a) enables the user to inspect and comprehend across the search system, and (b) the users' information needs become an essential entity for the search system in which the information need dynamics tend to articulate more to the system. An applicative instance of Ostensive retrieval can be rendered in a query reformulation process, where terms can be illustrated individually and turn into multi-term queries as information need evolves. The evolutive information needs in an ostensive model (OM) distinguish the grounded and evolving information needs in a given search process that supports a schematic model for rendering alteration. The ostensive model is used in the binary probabilistic model [66] for its assuredness, which circumvents the user interaction types - relevant and non-relevant to be captured, and so stipulating the user expression. Changes in information need are based on stipulated interactions that lead to a cognitive state change and thus act as a schematic model for it.

A user pivotal search can be rendered for OM and so is an Ostensive IR. Ostensive IR can be enhanced to characterise cognitive and psychological convictions by means of user interaction and their behaviour. This thesis views the ad-hoc nature of the ostensive model as an objective of the aforementioned explications, in particular, of modelling it as

one of the cognitive aspects. The principal aim is to standardise and formalise the ostensive model: (a) the cognitive model of dynamic information need, (b) the interaction model, (c) the informativeness of information need evolving from interaction, and (d) how it possesses retrieval. This thesis aims to develop an interactive framework using quantum probability formalism that addresses the above explications and the cognitive aspects with the help of Information Foraging theory which affects information need change. Primarily, it was the superiority of ostensive IR over traditional IR that inspires me to investigate via a formal viewpoint of ostensive IR.

Approaches in Interactive IR Computational search [27] refers to devising a query eliciting the users' information need situates varied character-based entities, for instance, an issue immanent to the query formulation task which concretely follows the characters in the process of satisfying their information needs at every state of occurring terms. This task in itself is a challenging problem for users' interpolation of their information needs which are mostly ambiguous but it transforms as they assess the search system. Such challenges give rise to an important aspect of a query being peculiar or complex and set forth a question on how the changing information needs are manifested and adapted by a search engine.

Furthermore, a distinguishable information need in which the majority of users find it hard to perceive the interaction among the user query and the system feedback. The formidable functionality of the retrieval system composes its efficiency and usefulness in order for the user to comprehend the system.

This led to the involvement of formal models - quantum-inspired IR, which is discussed in Section 2.3.6.

2.1.3 Reinforcement Learning for IR

Humans' transfer of information to other animals is a common method of learning and interaction, which is generally called reinforcement learning. Reinforcement learning [204] (RL) techniques are motivated by our sense of decision making in humans which appears to be biologically rooted. Within such biological roots [40], when an information forager's action ends up with a disadvantageous consequence (or negative payoff), such action will not be counted in the future; whereas, if his/her action leads to a successful consequence (or positive reward), it will happen again. User involvement in information searching is primarily a decision making (or action-taking) process [55], where users reflect identical RL features during this process. We will adopt RL models to manifest the mechanisms prevailing users' learning of information from searching. A typical scenario of the IR process in RL is shown in Figure 2.1. Previous work [229] found that a search system's information can

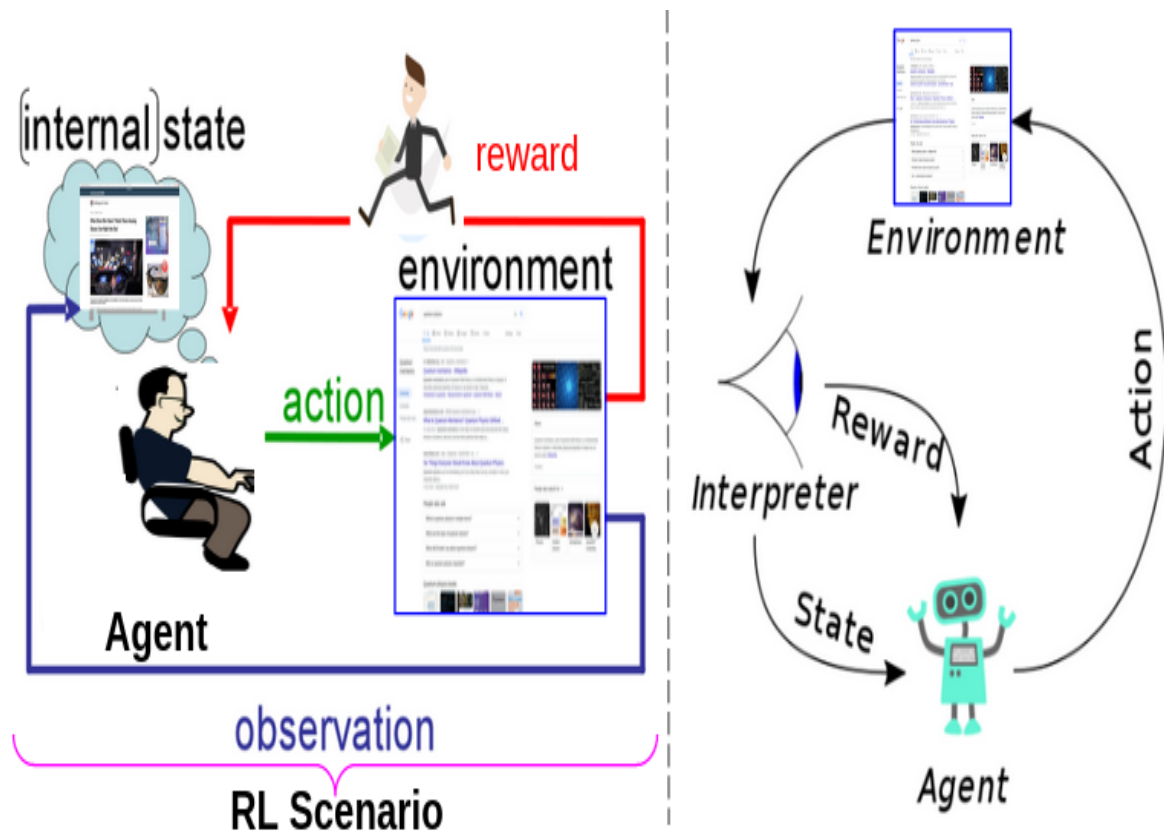


Fig. 2.1 IR process in a Reinforcement Learning Framework based on the general scenario

be enriched to advance search intention and automate the difficult query reformulation by modelling the search context. Reinforcement learning is an important method that can let the system employ the search context and relevance feedback simultaneously. Also, this approach allows the system to deal with exploration (widening the search among different topics) and exploitation (moving deeper into generic subtopics) which has been supportive in information retrieval [247, 192]. Exploration and exploitation methods are usually employed in tasks associated with recommender systems or information retrieval, such as foraging strategies [58], recommendation [244] or image retrieval [20]. However, reinforcement learning is mainly used by search/retrieval systems [255], which collect users' interests and habits over a continuous period, while in a specific search scenario the users in a given search session are more interested in the holistic improvement of the search results than relying on arbitrary future search sessions.

2.1.4 Query Auto-Completion

Users' search experience enhances via formulating their information needs (or queries). Initially, the commercial search engine Google Search provided a tool known as Google Suggest¹, which designates probable tail ends to the user queries once they type. This tool extends and is inbuilt under Google search as a mainstream provider to search completions in real-time. Search completion tools similar to Google Suggest follow the trend of searches analysed by users. [21] investigated this Google's tool and explicated as *query auto-completion* (QAC), and termed as most popular completion (MPC) that follows the search popularity of queries resembling the user's prefix. The gist of QAC can be perceived as a ranking task. For a prefix, there are probable query completions to it that ranks based on given criteria, and then a few of the optimal completions are reflected by the searcher.

There are certain types of QAC methods such as probabilistic [23, 159], user-oriented [93] and learning-based [196] models which broaden the capability of search completions depending on the required context.

We highlight briefly such QAC methods and their applicability in Chapter 5.

2.2 Introduction to Information Foraging Theory

Information Foraging theory is a theoretical framework to delineate the behaviour of information retrieval i.e., to characterise user search strategies and help them determine their information goal. It was first introduced by Peter Pirolli [166, 164] and the idea behind IFT is rooted in behavioural ecology, specifically, optimal foraging. Optimal foraging help animals determine their food search strategies which act as an initial heuristic behind the sound principles of Information Foraging. The heuristics of optimal foraging was investigated [7, 8] under a situational construct i.e., the environment in which a Web searcher forages to reach his/her information goal. However, the initial analysis [162] of comparisons among two different kinds of searchers i.e., an animal and a human user found to be similar in the perspective of their cognitive evolution while searching.

The notion of IFT is to shape the way of finding information for a searcher similar to an animal locating food. [164] founds that the evolving cognitive factors which help to find/gather information are an important characteristic to further foraging. One of the characteristics that optimal foraging possesses is exaptation [34]. This characteristic is an alternate mechanism of the forager nature which help them to narrow down their search to a particular environment that other forages. Similarly, the analogy of exaptation can be used

¹<http://labs.google.com/suggest>

for Information Foraging which let the user search behaviour adaptive to an information environment.

An information environment for a Web searcher is a place where several different types of information are distributed that human access across pertaining to certain objectives or inherently gathers the information. Toward the resemblance of IFT and optimal foraging [166, 162], the searching strategies differ only in terms of cognitive mechanisms through which an animal and a human search. This is because the information on the Web has its' unique way of representing information, whereas information in a physical environment is presented in a realistic form. A major component of Information Foraging is the utility behind accessing information. The utility here refers to the information gained pertaining to foraging (or searching). The principles of IFT act as a key driver in describing the searcher's cognitive aspects and their plethora of finding information for a certain information need. Earlier work [136] introduces commonalities between information and knowledge by means of IFT. They describe IFT as a complex activity of knowledge that contains locating needful information, and gathering and experiencing it. Such nature of Information Foraging paves the action of human behaviour for seeking.

Hence, IFT act as a sensemaking tool that evolves searching techniques over time in between the way of finding information and cooperating within an environment.

2.2.1 Key Constructs

IFT provides stipulated constructs adopted from optimal foraging theory which includes predators conforming to humans who seek information (or prey). They delineate these searches in the user interface sections, called patches. From the IFT perspective in image search, the searcher is the predator, the information patch is any segment or a region within an image in an artefact of the environment, the piece of information a user looks for is the prey, and the consumed (or gained) information is the information diet. A major aspect of something on the user's interface that informs a specific information object that they should look next refers to a *cue*. Information Foraging Theory constitutes three different models as an important construct to understand user behaviour:

Information Patch model deals with time allocation and information filtering activities.

Information Scent model helps people make use of perceptual cues, such as Web links spanning small snippets of graphics and text, consecutively to make their navigation decisions in selecting a specific link. The purpose of such cues is to characterise the contents that will be envisaged by trailing the links.

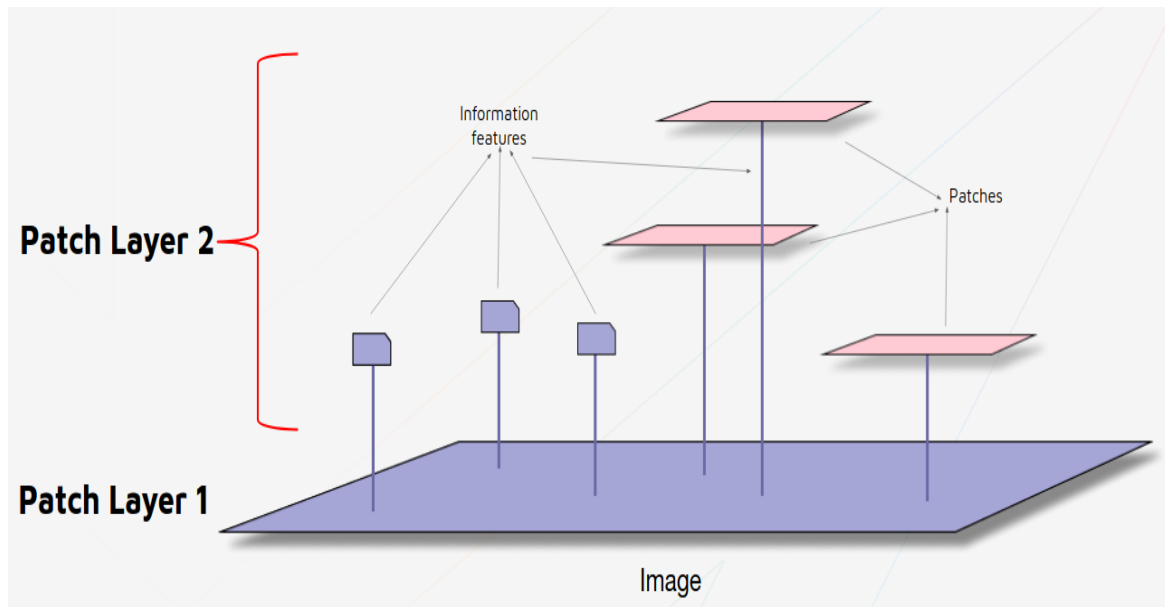


Fig. 2.2 Overview of Information Patch

Information Diet model deals with decisions about the combined set of information that has some perceived value to a searcher, who then pursues the set of information and neglects the remaining. Several types of information sources will vary in their prevalences or information access costs, emanates in information gain per unit cost, or varied profitability. The foragers will narrow or expand the diversities of information sources based on their profitability.

Information scent manifests the proximal cues and distal information sources they lead to (i.e. their perceived relevance). This theory infers that the user must seed their navigation decisions on the evaluation of information scent cues allied with individual choices. The user will evaluate the link likely to lead his/her information goal if the link cue perceived to have high information scent.

2.2.2 Role of IFT in IR

Information Foraging Theory [164] as a theoretical framework of information retrieval behaviour is an active topic of research in information science and human-computer interaction domain [84, 225, 189]. This section shed some insight into the usefulness of IFT in areas such as information retrieval and recommendation system.

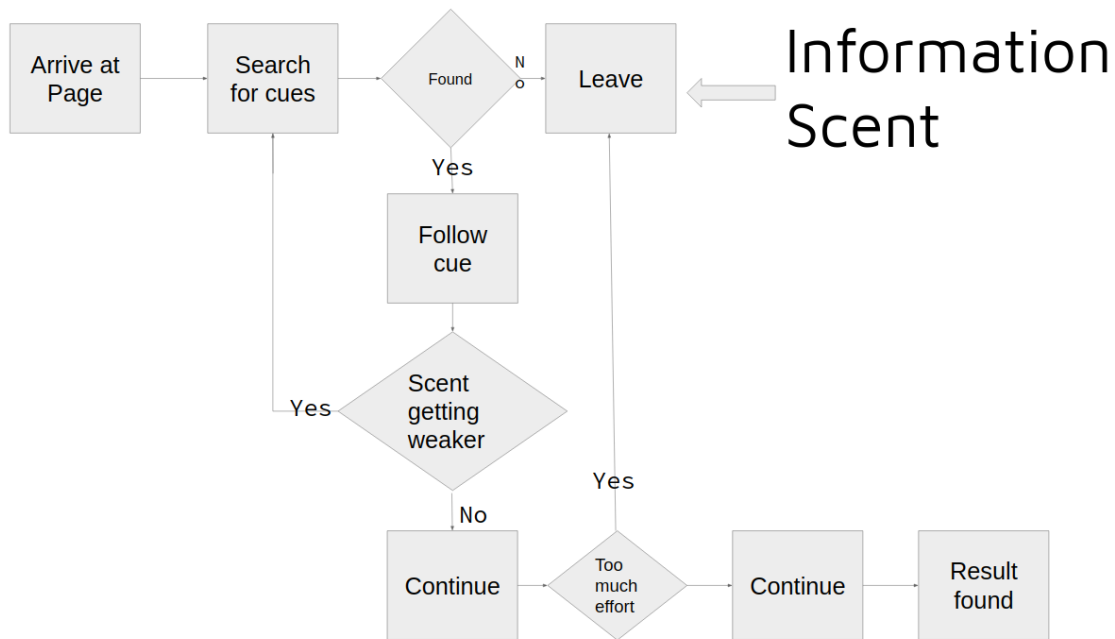


Fig. 2.3 Overview of Information Scent

Visual Information Foraging

Information visualisation [73, 167] is a key aspect of enriching human cognition in web search. The visualisation strategy's purpose is to emanate more information during the period of user attention. That makes the user consume more information in a short period of time, which increases the capacity of information processing, thus magnifying cognition. This motivates the recent work on search cost model [17] to minimise the cost framework of information. The first attempt on reducing the cost structure of information in the visual search was carried out in the context of Information Foraging by [167]. They assumed that the visual search is a complex system that evolves in various ways such as the density, number, kinds, and elements/items distribution in the visual field provided information presentation objective is to place relevant information on the user interface. They used IFT to support a hybrid framework that understands visual and cognitive search during interactions with visual interfaces to contents. Their framework leverages Hyperbolic Tree browser [190] to validate the notion of Visual Information Foraging. They found that their framework was dominated by information scent effects. Information scent also explicit user engagement in finding the concentrated area of the visual search engine that attracts user attention. The high and less concentration of users is estimated by strong and weak information scents. Their framework, named CODE Theory of Visual Attention (CTVA) based on Information Foraging was introduced to detect hierarchical sets of visual elements on a user interface

of visual search engines. They incorporated information scent to find which object can be foraged at low search cost with parallel search or high search cost with a serial search. They performed a qualitative analysis of user interface density, visual search, and information scent. This framework is extended to rectify users' capability to navigate large tree-structured information systems in [168]. [168] introduced Accuracy of Scent (AOS) to delineate a group of tasks that expect users to search or relate information in tree structures (search and relate means take part in retrieval and comparison tasks).

Information Foraging for Search and Recommendation Systems

Information Foraging Theory is matured enough to advance its applicability in web search (text search, image search, etc.) and recommendation systems. [189] use IFT to shape the user feedback data in movie recommendations, where user feedback allows the recommender system to infer the preferences based on information scent. They proposed a foraging-based strategy to investigate how the quality and quantity of feedback data can be shaped as well as learning according to alteration in the user interface provided the performance could be increased. They considered Information Foraging theory as a two-fold phenomenon: information scent and information access cost for understanding the feedback quality and quantity which can be impacted by interface design. User interface and the user enriches or learns information instantly when it starts interacting either by querying or recommending items. [189] rely on the data which was procreated from the implicit feedback when movie recommendation comes into play. They come up with an idea of *foraging intervention* which affects the generated implicit feedback data in two ways – the potency of information scent which is the information extent for an item shown at the beginning of recommendation, and the *information access cost* is the incurred cost for attaining more information about an item. Foraging intervention showed better performance in a movie recommendation task given the foraging strategy can effectively enrich the implicit feedback data. This work motivated researchers from two different communities: software engineering and machine learning to enjoin together to leverage the notion of Information Foraging theory. In image search, a recent work [68] on using Information Foraging theory for interactive image retrieval can be traced, where they enable information scent by providing search results with the help of visual cues to present the image collection and capability to estimate the outcomes of their interaction with the search interface. Personalised recommendation systems assist users in dealing with information overload by informing them about relevant items. When it comes to image recommendation, earlier studies [141] have relied on metadata and some work has leveraged visual features extracted with deep learning-driven neural networks to recommend art. However, the problem of information overload [202, 114] in an image

search (or recommendation) scenario has been studied by adjusting or enhancing the search interface [234] rather than by leveraging implicit feedback to improve recommender systems by learning user behaviour signals.

User Interaction

The process of a user's feedback or any interactive behaviour such as query reformulation in IR led to the exchange of information between the search system and the searcher. [169] made the first attempt to adopt Information Foraging to model user interaction in the web search scenario. They developed a *user-tracing* framework to simulate user interaction and relate the cognitive model of IFT i.e., SNIF-ACT [164] across user-trace data (or eye-tracking data). Their framework overlays observed user actions and relate with SNIF-ACT simulated actions. The user-tracing framework is an ecologically-inspired model based on Information Foraging theory. It builds on a task that looks like 'what people do in real, culturally significant situations' [149] and laboratory experiments are carried out to validate the task. The experiment record data using logging and eye tracker tools for user interactions within a web browser. The end goal of the user-trace model is to outcome with computational models of perception and user cognition.

Information Foraging theory has been applied to textual data to model users' information needs and their actions using information scent [42]. However, it has been previously found that information scent can analyse and predict the usability of a website by determining the website's scent [43]. Past work [129] demonstrated an IFT-inspired user classification model for a content-based image retrieval system to understand the users' interaction by functioning the model on several interaction features collected from the screen capture of different user tasks type. The user classification model constitutes of three factors: information goal (I), search strategy (S), and evaluation threshold (E). Each factor contains various types of user characters such as fixed information goal and evolving information goal, risky and cautious search strategy including the weak evaluation and precise evaluation threshold. The user study for standardising the ISE model captures 48 unique interaction features on different content-based image retrieval systems. They studied the usefulness of a foraging-based strategy in a content-based image search.

Recent work [96] studies the effects of foraging in personalised recommendation system by inspecting the visual attention mechanism to understand how users follow recommended items. A recent user study on modelling user interaction in web search is traced based on economic theory [14]. [14] developed an economic model of search to amplify the current interactive information retrieval systems based on the derived eight interaction-driven

conjectures in correspondence to search behaviour. The motive of this study is to suggest an explanation for observed or predicted search behaviours.

2.2.3 Language Model for Behavioural IR

The language model originates from the probabilistic IR, which computes the relevance probability for ranking documents. In this part of the Section, we interpret the language model based on the constructs of Information Foraging Theory. The interpretation is considered due to the fact that IFT explicates IR behaviour.

The information scent model, which takes into account the context of the user query, is constructed based on insights from language models. This model is characterised by a probabilistic framework that aligns more closely with the foundational principles of both the ACT-R and Information Foraging theories [164, 74].

The notion of IFT to derive the log-odds form of Bayes' rule for the information scent [163]. The terms can be explicated in a proximal cue [162] by 'd', where the proximal cue² is the title or authors or publisher of the document, and 'q' denotes the user's information need (or query). For estimating the information scent, it presumes the prior odds to be uniform given below:

$$\log\left(\frac{P(q|d)}{P(\bar{q}|d)}\right) = \log\left(\frac{P(q)}{P(\bar{q})}\right) + \log\left(\frac{P(d|q)}{P(d|\bar{q})}\right) \quad (2.5)$$

$$\log\left(\frac{P(q|d)}{P(\bar{q}|d)}\right) \approx \log\left(\frac{P(d|q)}{P(d|\bar{q})}\right) \quad (2.6)$$

where the component on left hand side of the Eq. 2.5 is the posterior log odds and $\log\left(\frac{P(q)}{P(\bar{q})}\right)$ is the prior odds of a given information need, and the component $\log\left(\frac{P(d|q)}{P(d|\bar{q})}\right)$ depicts the log likelihood ratio³ that a given information need is needed.

Information Foraging Theory posits that the prior probability of an information need, q , remains relatively stable both when the information need initially arises and when an item d is not expected. This assumption can be expressed mathematically as $P(d|\bar{q}) \approx P(q)$. To accommodate this assumption, the log-likelihood is modified to take the form $\log\left(\frac{P(d|q)}{P(q)}\right)$. The simplified log-likelihood is derived based on the aforementioned assumption.

²since this information is visible to the user and influences whether or not a document is clicked

³The log likelihood is the context-sensitive component and prior odds is independent of the context

$$\begin{aligned}
\log\left(\frac{P(d|q)}{P(q)}\right) &= \left(\frac{P(q \cap d)}{P(d)}\right) \left(\frac{1}{P(q)}\right) \\
&= \left(\frac{P(q)P(d|q)}{P(d)}\right) \left(\frac{1}{P(q)}\right) \\
&= \left(\frac{P(q|d)P(d)}{P(d)}\right) \left(\frac{1}{P(q)}\right) = \frac{P(q|d)}{P(q)} \quad (2.7)
\end{aligned}$$

where $P(q \cap d)$ is the conditional probability. This transformation presents the strength of association (S_{ji}) which measures how likely a source of evidence⁴ j is to be matched provided another concept i , and so is context-dependent, as reported in the equation below

$$S_{ji} \approx \log\left(\frac{P(i|j)}{P(i)}\right) \quad (2.8)$$

which exhibits the final activation A_i to calculate the information scent.

$$A_i = \sum_{j \in q} \log\left(\frac{P(i|j)}{P(i)}\right) \quad (2.9)$$

The ranking of documents can be elucidated as the Bayes' rule's likelihood component that asides an illustration to pointwise mutual information (PMI) whilst encompassing the normalisation of document length and word/term importance would be fictitious.

In information scent, the user information needs (query terms) are modelled as an action state (e.g. seeking documents in a given state (information need)) as perceived in the searcher's mind. Since, computing the information scent via smoothing based on the proximal cues (e.g. title of a document) to the information need. The cognitive modelling of user states where a state is contended as changing action (such as locating documents) that restrains information about the information need. It is the activation level (A_i) of the state that delineates the utility of the provided link (satisfying an information need) and the user clicks the links with strong information scent values (based on a high utility score).

Based on the above mathematical description, information scent can be formalised based on Dirichlet smoothing [246]. The element is denoted by e of the query q . Users assess additional information other than the title of the document can be traced by extra cues. The maximum likelihood estimate for the document language model computes the probability $P(e|d)$ based on the number of times e occurs in the proximal cue (d) and the frequency of

⁴seed term or term in second level reference

occurrence (e), all terms in the collection (c) respectively.

$$P(e|d) = \frac{P(e|d) + \mu P(e|c)}{|d| + \mu} \quad (2.10)$$

where μ is the pseudo count constant, which influences the amount of smoothing. Based on the Eq. 2.10, the language model with the demonstration from semantic relatedness score, which enables partial matching supports the notion of information scent. The aspect of information scent relies on a strong or weak signal of relevance that improves ranking performance. Most of the existing IR ranking models, such as BM25 and TF-IDF do not consider the role of context while querying or selecting documents. Researchers used information scent theory to understand web search behaviour [38, 240]. However, information scent can be a probable candidate to re-explore documents ranking. In [164], they introduced the cognitive framework of ACT-R (Adaptive Control of Thought-Rational), which advances that memory item with a prior probability distribution elicit the use of preceding memory for look-ahead needs. Provided a user information need (i.e., a query), the prior probabilities for the users' information needs are revised with the available information needs (presence of cues in the query), and the information need with the highest posterior probability is retrieved. The context sensitivity of the ACT-R theory of human memory emanates via the spreading activation equation, which incorporates the effect of the contexts provided by the cues. The association strengths (S_{ji}) between the concepts are recorded based on the past contexts in which these terms have appeared.

Based on the perspective of IFT, the user follow in selecting the hyperlink driven via proximal cues with the strong information scent, for instance, the Web Page (distal information) title to maximise the satisfying probability of the IN with the distal information content (e.g., the Web page allied with a hyperlink).

2.3 Desiderata of Quantum-inspired IR Models

This section elaborate on the fundamentals behind an elegantly formalised mathematical information retrieval (formal) framework, built upon the quantum mechanical framework. It also investigates how this mathematical description of quantum mechanics helps modelling the user's information need which spans from a given state to a choice via action space for understanding the user interaction process in the search scenario according to the outlook of the research prognosis in the following subsections. We have outlined the fundamentals of quantum mechanics with its mathematical description to quantum probability and its constructs in subsection 2.3.1 and subsection 2.3.2 on how to represent events geometrically

including some mathematical descriptions of quantum theory in Hilbert space [45] are presented with details on how the representation is functioning, then in accordance to these techniques for characterising user interaction are addressed.

2.3.1 Preliminaries

Vector Spaces

A vector space V with a given basis of \vec{v} is a (finite) set of specific vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ that are linearly independent. The vector components $v_i \in \{1, n\}$ can be real (or complex) numbers based on whether it lies in a real or complex vector space.

To represent a matrix vector space (M) and its basis (M_B), it can be delineated as

$$M_{2 \times 2} = \left\{ \begin{pmatrix} a & b \\ cd & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R} \right\}$$

$$M_B = \begin{pmatrix} a & b \\ cd & d \end{pmatrix} = a? + b? + c? + d?$$

The basis is a linear decomposition of the given matrix vector space.

A linear transformation in vector space is a mapping

$$T : V \rightarrow W$$

Where V and W are vector spaces such that

$$\forall \vec{u}, \vec{v} \in V, T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$$

and

$$\forall \vec{u} \in V, \forall \alpha \in \mathbb{R}, T(\alpha \vec{u}) = \alpha T(\vec{u})$$

A vector delineates a probability distribution and applying a matrix on the vector or a linear transformation reflects mapping probability distributions to probability distributions. This led to finding the associated probabilities of each vector via projection onto the basis vector.

$$\Pr(\vec{v}_1) = \vec{v}^T \cdot \vec{v}_1 = a$$

$$\Pr(\vec{v}_2) = \vec{v}^T \cdot \vec{v}_2 = b$$

where $\text{Pr}(\vec{v}_1)$ and $\text{Pr}(\vec{v}_2)$ represents probabilities of the \vec{v}_1 and \vec{v}_2 vector respectively.

Vector Space in Quantum Theory - Hilbert Space

A *column vector* in a complex vector space is written $|\psi\rangle$, and it refers to as a "ket",

$$|\psi\rangle = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

where $\alpha_1, \dots, \alpha_n \in \mathbb{C}$. Its *conjugate transpose* is written $\langle\psi|$, and its called a "bra".

$$\langle\psi| = |\psi\rangle^\dagger = (\alpha_1^* \dots \alpha_n^*)$$

A (finite-dimensional) *Hilbert space* [45], often denoted as \mathcal{H} , is a complex inner product space, i.e. a complex vector space equipped with a binary operator $\langle_|_ \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ called *inner product*, dot product, or simply "bra-ket".

$$\langle\psi|\phi\rangle = (\alpha_1^* \dots \alpha_n^*) \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \sum_i \alpha_i^* \beta_i$$

The inner product satisfies the following properties:

Conjugate symmetry	$\langle\psi \phi\rangle = \langle\phi \psi\rangle^*$
Linearity	$\langle\psi (\alpha \phi\rangle + \beta \phi\rangle) = \alpha\langle\psi \phi\rangle + \beta\langle\psi \phi\rangle$
Positive definiteness	$\langle\psi \psi\rangle \geq 0$

Notice that $\langle\psi|\psi\rangle = 0$ if and only if $|\phi\rangle$ is the $\mathbf{0}$ vector. Besides, thanks to conjugate symmetry, we have $\langle\psi|\psi\rangle = \langle\psi|\psi\rangle^*$, so $\langle\psi|\psi\rangle$ is always a real, non-negative number, when $|\phi\rangle \neq \mathbf{0}$.

Two vectors $|\psi\rangle$ and $|\phi\rangle$ are *orthogonal* if

$$\langle\psi|\phi\rangle = 0$$

The *norm* of $|\psi\rangle$ is defined as:

$$\| |\psi\rangle \| = \sqrt{\langle\psi|\psi\rangle}$$

A *unit vector* is a vector $|\phi\rangle$ such that

$$\| |\psi\rangle \| = 1$$

A set of vectors $\{|\psi\rangle_i\}_i$ is an *orthonormal basis* of \mathcal{H} if

- Each vector $|\phi\rangle \in \mathcal{H}$ can be expressed as a *linear combination* of the vector in the basis, $|\phi\rangle = \sum_i \alpha_i |\psi\rangle_i$.
- All the vector in the basis are orthogonal.
- All the vector in the basis are unit vector.

The basics introduced above are required for the postulates of Quantum Mechanics, as described below:

Postulate I: The state⁵ of a quantum system is represented, at a fixed time t , by a unit vector $|\psi\rangle$, called the *state vector*, belonging to a Hilbert space \mathcal{H} , also refers to the *state space*.

When describing the state of a quantum system, we ignore the *global phase factor*⁶, i.e.

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = - \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ for each } \lambda \in \mathbb{C} \text{ such that } |\lambda| = 1$$

The simplest, prototypical example of a quantum physical system is a *qubit*. In the aspect of quantum mechanics, a qubit (as opposed to the classical bit that encodes information in the form of 0 and 1) is associated with a two-dimensional complex Hilbert Space \mathcal{H}^2 . Such interpretations comprehend an electron in the ground or excited state, a vertically or horizontally polarised photon, or a spin-up or spin-down particle. It means that a vector or state informs one of the spin-up / spin-down components.

For instance, a photon, we could say that the photon is in state $|0\rangle$ when vertically polarised, and in state $|1\rangle$ when is horizontally polarised, where $|0\rangle$ and $|1\rangle$ are the two unit vector of the Hilbert space defined as

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

⁵The state refers to the physical state delineated by $|\rangle$

⁶An equivalent formulation, in fact, describes a quantum system as a *ray*, a one-dimensional subspace of \mathcal{H} .

The state vectors $\{|0\rangle, |1\rangle\}$ form an orthonormal basis, called the *basis vector*. Since they form a basis, each state vector $|\psi\rangle \in \mathcal{H}^2$ can be expressed as

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$.

So, to describe the state of a quantum system, any state vector can mathematically be described as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, a linear combination of $|0\rangle$ and $|1\rangle$. From the physical point of view, this means that this two-dimensional Hilbert space (\mathcal{H}^2) is in a *quantum superposition* of state $|0\rangle$ and $|1\rangle$, like a photon being diagonally-polarised, or an electron being at the same time in the excited and the ground state.

Other important vectors in the \mathcal{H}^2 state space are $|+\rangle$ and $|-\rangle$,

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \\ |-\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle \end{aligned}$$

that form the so-called *basis* of \mathcal{H}^2 . As we will see, $|+\rangle$ and $|-\rangle$ are both an equal superposition of $|0\rangle$ and $|1\rangle$. The difference between $|+\rangle$ and $|-\rangle$ is the *relative phase*, i.e. the phase between the $|0\rangle$ and $|1\rangle$ component. The states $|0\rangle + |1\rangle$, $-|0\rangle - |1\rangle$, $i|0\rangle + i|1\rangle$ are all equal to $|+\rangle$ times a certain global phase and are considered the same state. The states $|0\rangle + |1\rangle$, $|0\rangle - |1\rangle$, $|0\rangle + i|1\rangle$, instead, all differ for a relative phase factor and have different behaviours when applied to the same computation.

Unitary Transformations

For each linear operator A acting on a Hilbert space \mathcal{H} , we denote as A^\dagger the *adjoint* of A , i.e. the unique linear operator such that

$$\langle \psi | A \phi \rangle = \langle A^\dagger \psi | \phi \rangle$$

A linear operator A acting on a n -dimensional Hilbert space \mathcal{H}^n can be represented as a $n \times n$ matrix, and its adjoint is calculated as

$$A^\dagger = (A^*)^T$$

the conjugate transpose of the matrix A .

If it holds that $A = A^\dagger$, we say that A is self-adjoint, or *Hermitian*.

A linear operator U is said to be *unitary* when $U^\dagger = U^{-1}$, which implies

$$UU^\dagger = U^\dagger U = I$$

Unitary matrices enjoy many useful properties, first of all, that they have a spectral decomposition. An other defining characteristic is that they preserve the inner product, $\langle \psi | \phi \rangle = \langle U\psi | U\phi \rangle$

$$\langle U\psi | U\phi \rangle = \langle \psi | U^\dagger U | \psi \rangle = \langle \psi | I | \psi \rangle = \langle \psi | \phi \rangle$$

A corollary of this property is that applying a unitary operator to a unit vector gives a unit vector

$$\langle U\psi | U\psi \rangle = \langle \psi | \psi \rangle = 1$$

The following postulate makes obvious why we are interested in unitary transformation.

Postulate II: The evolution of a closed quantum system is described by a unitary transformation. That is, the state $|\psi\rangle$ of the system at time t_0 is related to the state $|\psi'\rangle$ of the system t_2 by a unitary operator U which depends only on the times t_0 and t_1 .

$$|\psi'\rangle = U |\psi\rangle$$

According to the aforementioned description of the unitary transformation, if a physical system starts in a unit state, it will always remain in a unit state. For instance, a linear transformation over a unitary Hilbert space (or a vector space over the complex field)⁷.

Measurement

The second postulate describes only the evolution of quantum systems. Such systems do not exchange information with the environment, and all their computations are always reversible (and are in fact formalised with invertible, unitary matrices). To extract classical information from the system to *measure* the output of the quantum computation, an interaction is needed between the quantum system and the environment. As we will see, this measurement operation is first of all non-invertible, as a different state could produce the same outcome when measured, but is also fundamentally probabilistic. A generic state $|\psi\rangle$ could produce

⁷https://encyclopediaofmath.org/wiki/Unitary_transformation

different measurement outcomes m_1, m_2, \dots , each with a certain probability that depends on $|\psi\rangle$.

From a physical point of view, if a system is in a superposition of states, measuring it can cause the wave function to collapse to a single state, in a purely probabilistic way. This means that, even if we compute a state that contains the desired information, this information is often difficult to recover, because directly measuring it can destroy the information and produce a trivial outcome.

Postulate III: Quantum measurements are described by a set $\{M_m\}_m$ of measurement operators, where the index m refers to the measurement outcomes that may occur in the experiment. The set of measurement operators must be *complete*, i.e.:

$$\sum_m M_m^\dagger M_m = I$$

If the state of the quantum system is $|\psi\rangle$ before the measurement, then the probability that results from m occurs is

$$p_m = \langle \psi | M_m^\dagger M_m | \psi \rangle$$

and the state after the measurement will be

$$\frac{1}{\sqrt{p_m}} M_m |\psi\rangle$$

The most common class of quantum measurements is composed of *projective measurements*. Such measurements are described by a set of *orthogonal projectors*, i.e. Hermitian operators such that

$$M_m M_{m'} = \begin{cases} \mathbf{0} & \text{if } m \neq m' \\ M_m & \text{if } m = m' \end{cases}$$

The simplest example of (projective) measurement is simply measuring a state in the basis, i.e. projecting it in its 0-1 component. The measurement in the basis denoted as M_{01} is defined as

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad M_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

And its effect of the state $|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ is:

$$\frac{1}{\sqrt{p_0}}M_0|\psi\rangle = \frac{1}{|\alpha|} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle \quad \text{with probability } \langle\psi|M_0^\dagger M_0|\psi\rangle = |\alpha|^2$$

$$\frac{1}{\sqrt{p_1}}M_1|\psi\rangle = \frac{1}{|\beta|} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |1\rangle \quad \text{with probability } \langle\psi|M_1^\dagger M_1|\psi\rangle = |\beta|^2$$

Notice that, when measuring the $|0\rangle$ state in the basis, the outcome will always be $|0\rangle$ with probability $|\alpha|^2 = 1$, in a completely deterministic behaviour. When instead measuring the $|+\rangle$ state we get either $|0\rangle$ or $|1\rangle$, with equal probability $|\alpha|^2 = |\beta|^2 = \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{2}$. The same holds also for the $|-\rangle$ state. As described earlier, $|+\rangle$ and $|-\rangle$ differ only for the relative phase, and the relative phase has no influence on the outcome of measurement in the basis vector.

Another projective measurement is a measurement in the basis vector, M_\pm , defined as

$$M_+ = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad M_- = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

From the physical point of view, this measurement can be performed by composing the basis vector transformations and measurement in the basis: $M_+ = HM_0H$, $M_- = HM_1H$. Just as M_{01} makes a vector decay in its $|0\rangle$ component or in its $|1\rangle$ component, M_\pm makes a vector decay in its $|+\rangle$ or $|-\rangle$ component. This means that, when measuring $|+\rangle$ in the basis vector, the outcome will always be $|+\rangle$ with probability 1. However, a quantum system in the state $|0\rangle$ can be considered in an equal superposition of $|+\rangle$ and $|-\rangle$, as $|0\rangle = \frac{1}{\sqrt{2}}|+\rangle + \frac{1}{\sqrt{2}}|-\rangle$. According to this, measuring $|0\rangle$ in the basis gives the outcomes $|+\rangle$ or $|-\rangle$ with half the probability each.

Basics of Tensor Product

In the previous sections, we characterised quantum systems and how they evolve, taking the prototypical example of the two-dimensional Hilbert space. When dealing with higher-dimensional systems, we can often describe them as *composite system*, made of multiple elementary quantum systems.

If for instance, we have two quantum systems, each described by a (2-dimensional) Hilbert space \mathcal{H} encoding their corresponding states in it. We can describe the system composed as the *tensor product* (denoted by \otimes) of the 2-dimensional Hilbert spaces.

If \mathcal{H}_n is a n -dimensional Hilbert space, and \mathcal{H}_m is a m dimensional Hilbert space, their tensor product $\mathcal{H}_n \otimes \mathcal{H}_m$ is a nm Hilbert space. If $\{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ is a basis of \mathcal{H}_n , and $\{|\phi_1\rangle, \dots, |\phi_m\rangle\}$ is a base of \mathcal{H}_m , then a basis of $\mathcal{H}_n \otimes \mathcal{H}_m$ is

$$\{|\psi_i\rangle \otimes |\phi_j\rangle, \forall i \in [1, \dots, n], j \in [1, \dots, m]\}$$

where $|\psi\rangle \otimes |\phi\rangle$ denotes the Kronecker product. For convenience, certain authors omit the tensor symbol, writing $|\psi\rangle|\phi\rangle$ or also $|\psi\phi\rangle$ instead of $|\psi\rangle \otimes |\phi\rangle$. Below is the required postulate:

Postulate IV: The state space of a composite physical system is the tensor product of the state spaces of the component physical systems.

If a single quantum system is described by a 2-dimensional space \mathcal{H} , it can be described as $\mathcal{H}^{\otimes n}$ to intend the \otimes product n copies of \mathcal{H} of dimension 2^n . So, a compound system composed of two quantum systems has a state space $\mathcal{H}^{\otimes 2}$, its canonical basis is

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$$

and all its vectors can be expressed as a linear combination:

$$|\psi\rangle \in \mathcal{H}^{\otimes 2} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = \alpha |00\rangle + \beta |01\rangle + \gamma |10\rangle + \delta |11\rangle$$

A quantum state in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is said *separable* when can be expressed as the product of two vectors, one in \mathcal{H}_1 and the other in \mathcal{H}_2 . From the definition of the Kronecker product, all separable states of $\mathcal{H}^{\otimes 2}$ are of the form:

$$|\psi\rangle = \begin{pmatrix} \alpha\gamma \\ \alpha\delta \\ \beta\gamma \\ \beta\delta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \otimes \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$$

Some of the defining characteristics of quantum systems derive from the fact that not all states in $\mathcal{H}^{\otimes 2}$ are separable. The existence of such states, called *entangled* states, implies that

a composite system can not always be described as simply the juxtaposition of two smaller states. When a quantum system q_1 is entangled with another quantum system q_2 , its evolution depends not only on the transformations applied to q_1 , but also on the transformations applied to q_2 .

The classical example of an entangled state is the so-called $|\Phi^+\rangle$ *Bell state*:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

The fourth postulate tells us that the state space of a composite system is simply the tensor product of the state spaces of the smaller systems. The tensor product of Hilbert spaces is still a Hilbert space, the composition of unit vectors is still a unit vector, and the composition of unitary transformation is still unitary. For example, the (Kronecker) composition of H and I matrices is defined as a block matrix:

$$H \otimes I = \frac{1}{\sqrt{2}} \begin{pmatrix} I & I \\ I & -I \end{pmatrix}$$

And when applied to a two-quantum system, it applies H on the first quantum system, and leaves the second one unaltered:

$$(H \otimes I)|00\rangle = H|0\rangle \otimes I|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} I & I \\ I & -I \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = |+\rangle \otimes |0\rangle$$

Density Operator

The formalism and constructs represented so far describe the quantum system in terms of unit vectors and linear transformations. There is an alternative, more general formulation, the *density operator*, in which states are represented as positive operators, and transformations as linear maps from operators to operators, i.e. superoperators.

The main advantage of this formulation is that it represents also *partial information* about a quantum system. When describing quantum systems that interact with the external environment, it is often impossible to have complete knowledge of the state of our systems. Instead, one could know that a given system is either in state $|\psi\rangle$, with a certain probability

p , or in state $|\phi\rangle$, with probability $1 - p$. In other words, we know that the system is in a *probabilistic mixture of states*, called an *ensemble* of states, or also a *mixed state*.

In general, given an n -dimensional Hilbert space \mathcal{H} , an *ensemble* of quantum states is a set:

$$\{(|\psi_i\rangle, p_i)\}$$

of quantum states in \mathcal{H} , each with a different probability, such that $\forall i p_i > 0$ and $\sum_i p_i \leq 1$. Notice that when $\sum_i p_i = 1$, they have a probability distribution of states, when $\sum_i p_i < 1$, they have a so called subprobability distribution.

Each ensemble defines a density operator, that is a density matrix in $\mathbb{C}^{n \times n}$, i.e. an operator $\mathcal{H} \rightarrow \mathcal{H}$. The ensemble $\{(|\psi_i\rangle, p_i)\}$, where $|\psi_i\rangle \in \mathcal{H}$, defines the density operator:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

where $|\psi\rangle\langle\phi|$ denotes the matrix product between the column vector $|\psi\rangle$ and the row vector $\langle\phi|$, known as the *outer product*. Notice that this construction is not injective, as there are different ensembles that correspond to the same density operator. We indicate with $\mathcal{D}(\mathcal{H})$ the set of density operators of \mathcal{H} .

Density operators enjoy two useful properties (see [151]):

1. The trace of ρ is the sum of the probabilities of the ensemble $tr(\rho) = \sum_i p_i \leq 1$.
2. ρ is *positive semidefinite*, i.e.

$$\forall |\psi\rangle \in \mathcal{H} \quad \langle\psi|\rho|\psi\rangle \geq 0$$

Positive semidefinite operators are always diagonalisable with eigenvalues real and positives. So, each positive semidefinite operator with trace ≤ 1 represents at least one ensemble, with the eigenvectors as states and the corresponding eigenvalues as probabilities.

One of the key applications of density operators is to describe the state of a subsystem of a composite quantum system. When dealing with composite system, we write $\rho \otimes \sigma \in \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, to denote the density matrix given by the Kronecker product of ρ and σ . Notice that, if $\rho = \sum_{i=1}^n p_i |\psi_i\rangle\langle\psi_i|$ and $\sigma = \sum_{j=1}^m p_j |\phi_j\rangle\langle\phi_j|$, then

$$\rho \otimes \sigma = \sum_{i=1}^n \sum_{j=1}^m p_i p_j |\psi_i \phi_j\rangle\langle\psi_i \phi_j|$$

Suppose a composite system, made of two subsystem A and B , with state space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Given a generic (not necessarily separable) $\rho^{AB} \in \mathcal{H}$, describing the state of the

whole system, the operator ρ^A that describes the subsystem A is obtained as

$$\rho^A = \text{tr}_B(\rho^{AB})$$

where the tr_B is called the *partial trace over B* and is defined by

$$\text{tr}_B(|\psi\rangle\langle\psi'| \otimes |\phi\rangle\langle\phi'|) = |\psi\rangle\langle\psi'| \text{tr}(|\phi\rangle\langle\phi'|)$$

together with linearity. ρ^A is called the *reduced density operator* of system A .

When applied to a separable state $\rho^A \otimes \rho^B$, the partial trace tr_B gives exactly ρ^A . When applied to an entangled state, instead, it produces a probabilistic mixture of states, because "forgetting" the information on the state of subsystem B leaves us with only partial information on subsystem A . The canonical example is the bell state $\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11| \in \mathcal{H}_A \otimes \mathcal{H}_B$:

$$\begin{aligned} \text{tr}_B(\rho) &= \frac{1}{2}\text{tr}_B(|00\rangle\langle 00|) + \frac{1}{2}\text{tr}_B(|11\rangle\langle 11|) \\ &= \frac{1}{2}\text{tr}_B(|0\rangle\langle 0| \otimes |0\rangle\langle 0|) + \frac{1}{2}\text{tr}_B(|1\rangle\langle 1| \otimes |1\rangle\langle 1|) \\ &= \frac{1}{2}|0\rangle\langle 0| \text{tr}(|0\rangle\langle 0|) + \frac{1}{2}|1\rangle\langle 1| \text{tr}(|1\rangle\langle 1|) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}I \end{aligned}$$

The state $\frac{1}{2}I$ is called the *maximally mixed state*, as it gives us no information on the system. It can be seen as an ensemble of states $|0\rangle$ and $|1\rangle$ both with probability one half, but could also indicate an ensemble of $|+\rangle$ and $|-\rangle$ with the same probability, or even an ensemble of $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$, and so on.

Using density matrices and superoperators, it is possible to restate all the quantum postulates detailed in the previous paragraphs. A physical system is represented by a density matrix $\rho \in \mathcal{D}(\mathcal{H})$ with trace 1. A detailed description is given below:

Let \mathcal{H}_1 and \mathcal{H}_2 be Hilbert spaces. Given $\mathcal{E} : \mathcal{D}(\mathcal{H}_1) \rightarrow \mathcal{D}(\mathcal{H}_1)$ and $\mathcal{F} : \mathcal{D}(\mathcal{H}_2) \rightarrow \mathcal{D}(\mathcal{H}_2)$, their tensor product $\mathcal{E} \otimes \mathcal{F} : \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2) \rightarrow \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is defined as:

$$(\mathcal{E} \otimes \mathcal{F})(\rho_1 \otimes \rho_2) = \mathcal{E}(\rho_1) \otimes \mathcal{F}(\rho_2)$$

together with linearity.

Two quantum systems each described in their own Hilbert spaces is described by a density matrix $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. That is, the state ρ of the quantum system at time t_0 is related to the state ρ' of the quantum system at time t_2 by a superoperator \mathcal{E} which depends only on

the times t_0 and t_1 .

$$\rho' = \mathcal{E}(\rho)$$

Superoperators can describe both unitary transformations and measurements, as well as more general transformations. Given unitary operator U on \mathcal{H} , one can define the (trace-preserving) superoperator $\mathcal{E}_U \in \mathcal{S}(\mathcal{H})$ as:

$$\mathcal{E}_U(\rho) = U\rho U^\dagger$$

Given a measurement $\{M_m\}_m$, one can define the (trace non-increasing) superoperator $\mathcal{E}_m \in \mathcal{S}(\mathcal{H})$ as

$$\mathcal{E}_m(\rho) = M_m\rho M_m^\dagger$$

Here $\mathcal{E}_m(\rho)$ is equal to $p_m\rho_m$, where p_m is the probability of the outcome m when measuring the state ρ , and ρ_m is the state after outcome m has occurred.

2.3.2 Introduction to Quantum Probability

In quantum theory, each event is represented as multi-dimensional vectors in a Hilbert space [45, 256]. The vector form of representation manifests the potential of all events. In contrast to quantum mechanics, this property introduces the principle of superposition. Based on the psychological point of view, a quantum superposition can be concerned with the notion of confusion, uncertainty, or ambiguity [33]. It also facilitates the creation of more extensive models that can mathematically elucidate cognitive phenomena geometrically [140, 171, 90, 91], which is the key focus of this thesis.

This section presents the connection of the fundamental concepts introduced in subsection 2.3.2 to the quantum-inspired IR framework that is essential for the understanding of this work.

Quantum State In quantum theory, the appropriate notation is the Dirac notation (also known as "bra-ket" notation), where column vectors are written in a linear and dense way. For example, an l -dimensional column vector A :

$$A = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{l-1} \end{pmatrix}$$

can be expressed in terms of *ket* notation, as $|A\rangle$, such as:

$$|A\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \dots + \alpha_{l-1} |l-1\rangle$$

where,

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \dots, |l-1\rangle = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

Thus, the inner product of vector A can be expressed in terms of Dirac notation as:

$$\langle A|A\rangle = \left(\alpha_0, \alpha_1, \dots, \alpha_{l-1} \right) \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{l-1} \end{pmatrix} = |\alpha_0|^2 + |\alpha_1|^2 + \dots + |\alpha_{l-1}|^2.$$

where $\langle A|^\dagger = |A\rangle$

which is known as the quantum projection operator.

Using the above equations, a quantum state can be written as an extension into a complete orthonormal system

$$|A\rangle = \sum_i c_i |n\rangle \quad (2.11)$$

where the coefficients $c_i = \langle n|A\rangle$ are complex numbers representing the projections of the state $|A\rangle$ onto the eigenstates $|n\rangle$.

The key advantage of dealing with Dirac's notation is that it allows the explicit labelling of the basis vectors.

Hilbert Space Generally, the introduction of the wave function is presented through its conventional representation, as a vector in a Hilbert space \mathcal{H} for QT [45, 52]. The wave function is an illustration of the state of a physical system (or observable) at a particular point in time. A Hilbert space symbolised by \mathcal{H} is a vector space with the norm and an inner product such that it is a complex metric space. In quantum theory, the representation space is a Hilbert space on a complex-valued field, in which a vector's coordinates are elaborated

by complex numbers. Conventionally Dirac notation eases the process of depicting vectors and operations on this space, the notation in actuality reduces the visual complexity of demonstrating the intended mathematics, elucidating and minimising the effort of the overall calculation.

In quantum probability theory [52, 256], events are accommodated in a Hilbert Space. A Hilbert space can be described as an extension and generalisation of the Euclidean space into linear vector spaces with any finite or infinite number of dimensions. The vector space is traversed by a set of orthonormal basis vectors, which form a basis for the Hilbert space. For instance, in jury service example, we can define a set of orthonormal basis vectors as $\mathcal{H} = \{|\text{Offender}\rangle, |\text{Guiltless}\rangle\}$, where

$$|\text{Offender}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |\text{Guiltless}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Figure 2.4, presents a diagram illustrating the Hilbert space of a defendant being an offender or guiltless [33]. Through a Hilbert space validates the practice of complex numbers (it refers to the fact that a Hilbert space follows the properties of the inner product [10] in quantum mechanics which is why it is also referred to as a complex inner product space), then, systematically to describe the events Offender and Guiltless, one would specify two dimensions for each event (i.e., for the real part and another for the imaginary part). In quantum decision theory, one often set-asides the imaginary component appropriately to be able to visualise geometrically all vectors in a two-dimensional space.

Pertaining to the jury service example, the superposition state $|S\rangle$ can be defined as a superposition of an indictee being both offender and guiltless, reported in Figure 1 which presents a geometric visualisation of event $|S\rangle$ in a Hilbert space.

$$|S\rangle = \frac{e^{i\theta_O}}{\sqrt{2}} |\text{Offender}\rangle + \frac{e^{i\theta_G}}{\sqrt{2}} |\text{Guiltless}\rangle$$

where the factors $\frac{e^{i\theta}}{\sqrt{2}}$ are quantum probability amplitudes.

If we presume the condition where the jury needs to judgment about the guiltiness of two indictees, who were condemned for leading a crime together, then one could write their joined state using an operation called *tensor product*, which is depicted by the symbol \otimes . The tensor product is associated with a mathematical method that entitles the construction of a Hilbert space from the conjunction of individual spaces. If we have two indictees who could be each either Offender or Guiltless, then we could characterise them as a complex linear combination of these states as:

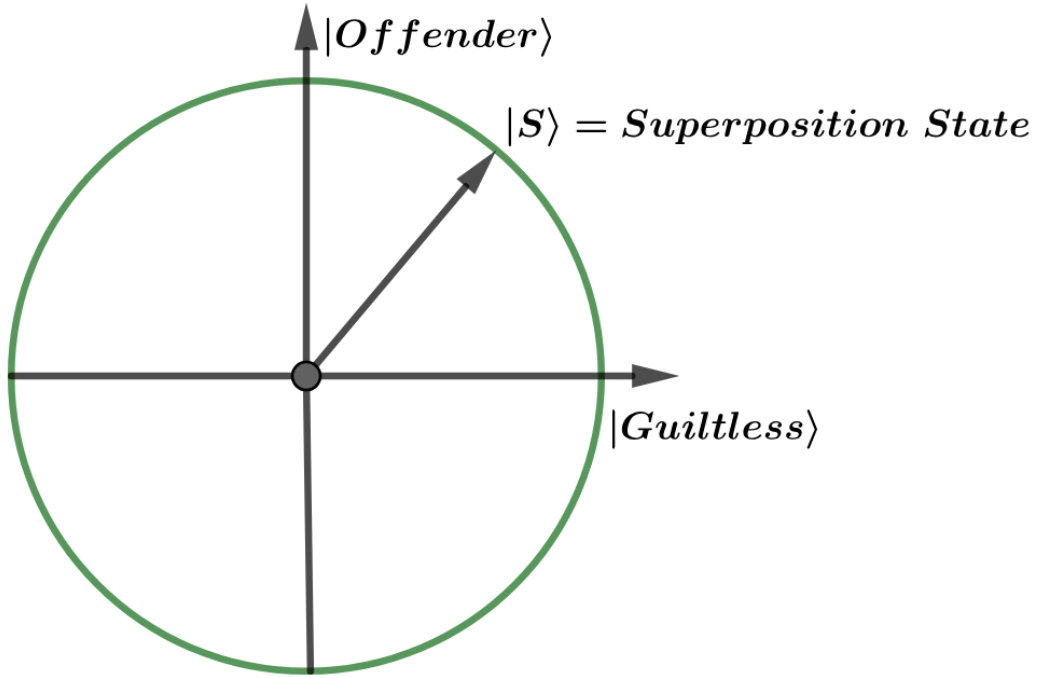


Fig. 2.4 Hilbert Space

$$(\alpha_0 |O\rangle + \beta_0 |G\rangle) \otimes (\alpha_1 |O\rangle + \beta_1 |G\rangle) = \alpha_0\beta_0 |OO\rangle + \alpha_0\beta_1 |OG\rangle + \alpha_1\beta_0 |GO\rangle + \alpha_1\beta_1 |GG\rangle$$

where O and I, depict the basis for Offender and Guiltless, respectively. And α_i, β_j , signify the quantum complex amplitudes connected to the first and second indictees, respectively.

Anatomy of Quantum Probability

The notion of quantum theory is inspired by the view of an IR system as analogous to a quantum system used in physics. In this section, we overview the mathematical formalism [211] behind the quantum-theoretic models [170, 64] of IR. In contrast to set-based classical probability theory, in quantum probabilities, the probabilistic space is geometrically defined, and its representation becomes an infinite set of angles and distances in a finite or infinite-dimensional Hilbert Space [201] denoted by \mathcal{H} . Each and every event is a subspace of the Hilbert Space. To represent the n -dimensional vectors that compose a Hilbert Space, the Dirac notation is widely adopted, using *ket* and *bra* nomenclatures. More concretely, this means representing one given vector ψ as $|\psi\rangle$ and its transposed view, ψ^T as $\langle\psi|$. Also,

the vectors under consideration in a Hilbert space are called *state vectors*, which are unit vectors. The squared length of the orthogonal projection of a state vector onto subspaces representing events (denoted by the projector $|\psi\rangle\langle\psi|$) induces a probability distribution over these quantum events. In a quantum system, there can be more than one state vector; a probability distribution over state vectors (which is different from the probability distributions induced by every state vector) reflects the uncertainty about the state the system might be in. The unit vectors-induced probability distributions over events (subspaces), can be represented by a so-called *density matrix*, denoted by ρ . Each state vector is a linear combination of the eigenvectors spanning the subspace. Each quantum event has its quantum probabilities defined by density matrices. In this thesis, we deal with a composite system that constitutes two Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 corresponding to textual query space and image representation space. We join these two individual systems from two-subspaces by means of the Tensor product of two Hilbert spaces. We assume that the dimensionality of the tensor space ($\mathcal{H}_1 \otimes \mathcal{H}_2$) is finite. Let $|h_{1_i}\rangle : i = 1, 2, \dots, n$ form an orthogonal basis for \mathcal{H}_1 and similarly the orthogonal basis for \mathcal{H}_2 is $|h_{2_j}\rangle : j = 1, 2, \dots, m$. Therefore, $|h_{1_i}\rangle \otimes |h_{2_j}\rangle$ forms the basis of the tensor space ($\mathcal{H}_1 \otimes \mathcal{H}_2$).

2.3.3 Probabilistic Nature in Quantum IR Models

The pioneering model of IR by [140] deals with general quantum probability to model context and suggests a probabilistic retrieval function to itemise the probability of relevance in the objects presented in a vector space. In general, classical systems such as search engines are unfamiliar with such an extremely dynamic search environment and contextual characteristics are not utilised during retrieval nor are they captivated at indexing time. Thus, retrieval may be imprecise, specifically when one-word queries are provided to the system; it needs to be disambiguated since they are contrived while not explicitly referring to the features provided by users' previous searches, backgrounds, and objectives.

In this IR framework, they use the concept of the *basis* of a vector space [139] which is orthonormal and signifies a contextual property such as the location, space, time, or sense of an information object. To be exact, any vector of a subspace is a linear combination of a basis for that subspace. It has been illustrated that the probability of context i.e., the probability that an information object remains materialised⁸ inside a context. This model is convenient for developing a classical information system in which relevant documents to information need to correspond against a query without radically considering context into account. Although, they employ a basis as the data structure for characterising information

⁸The term *materialisation* signals process of making information substantial as data

about the query and relevance, where the query acts as a subspace, and relevance poses the inclination of the subspace and the distance from those subspaces to represent the documents. The model is defined over the complex field, but the properties of density matrices, projectors, and trace functions persuade that the probabilities will be real and has their own limitation when it comes to evaluation, especially evaluating IR in context systems.

2.3.4 Generalisation of Language Embeddings

Several information retrieval or natural language processing tasks [135, 120, 61, 77, 88, 59] use different types of language embeddings, such as word2vec [23, 143], Glove⁹, and fastText¹⁰ which construct semantic representations of words based on their contexts. Their applicability and usefulness have been significantly demonstrated in different low-level tasks such as document ranking, entity associations, text classification, sense modelling, word associations, etc. The basic idea of word embedding is to employ word co-occurrence as the basis of the semantic relationship among words. This leads to a challenge for current embedding models that fail to capture the contingent meaning of words in combination, such as a sentence or a phrase. To overcome this problem in embedding models, [124] adopted the idea of quantum theory with the usage of complex numbers to incorporate the polarity of words. Their model – complex word embedding has been successfully applied to a variety of tasks such as text classification, text matching, etc. Their embedding model is inspired by the famous double-slit experiment where they modelled the composite words in mixed state (or superposition state) as quantum states having both amplitudes and complex phases. The limitation of this embedding model is due to resource-intensive. We introduce a computationally optimised method for the complex embedding of words [95] to overcome this problem. The challenge of capturing the word meaning has been extensively explored and extended to sentence-level [216], they considered the notion of quantum cognition which imply that human cognition [33] specifically language understanding [32, 218] reveals some non-classical occurrence. They introduced a quantum-like theoretical framework for language understanding – Semantic Hilbert Space, which models various levels (the non-linear semantic composition of words) of semantic units in a unified complex-valued vector space. The feasibility of their framework has been validated in text classification tasks with various benchmark datasets. Additionally, this is one of the first works in text modelling via quantum theory and quantum probability. They further extended the Semantic Hilbert Space model in matching tasks [125] with a focus on solving an interpretable matching problem, where they incorporate all the linguistic components in a unified complex-valued Hilbert

⁹<https://nlp.stanford.edu/projects/glove/>

¹⁰<https://fasttext.cc>

space with mathematical descriptions and definite physical meaning. The complex-valued matching network [125] showed its' effectiveness on question-answering tasks with better interpretability and explanatory power.

2.3.5 Generalised Classical Language Models

Language model based on the quantum probabilistic framework was first introduced in [201] to model dependencies between terms, referring to Quantum Language Model (QLM). The mathematical formalism of quantum theory models the complex dynamics and interactions in quantum physics and has been embraced for generating text representations in multiple information retrieval and natural language processing (NLP) tasks [201, 124, 253, 241, 250]. The principled quantum language model [201] describes a document or query as a density matrix in a quantum probabilistic space and estimates density matrix-based measures as a ranking function. [250] develops an end-to-end neural network-based quantum language model for question answering to compositely model a question-answer pair by using the density matrix representation. Additionally, [253] adopted QLM to extend the base model for sentiment analysis tasks on the Twitter dataset. These quantum language models can be regarded as a generalised version of classical perspectives in those IR and NLP tasks, and are competent in capturing implicit involutions in interactions. Recent work on the adoption of quantum theory in conversational sentiment analysis [252] is proposed to model the intra-/inter-utterance interaction dynamics which altogether is a complex task. They introduced a quantum-like interactive network that exploits both quantum theory and long short-term memory network to learn the shared interaction. Their model performs feature extraction on textual data with the usage of density matrix-based convolutional neural network (CNN) to inherently capture the inter-utterance correlations among words. The inter-utterance interaction mechanism is driven by quantum measurement theory which measures the impact among speakers across utterances. The notion of this quantum-like interactive network is to effectuate the consequence of modelling interactions.

2.3.6 User Interactions

Polyrepresentation is a theoretical framework of cognitive information retrieval [90]. It refers to cognitive overlaps as the structurally different representation of a document generated by functional and divergent cognitive representations. The polyrepresentative framework of IR aims to satisfy users' information needs by combining different representations of a document or query [63, 57] in Hilbert space. [63] introduced the principle of Polyrepresentation in a quantum-like IR framework to represent a document by combining various representations

of its' metadata via tensor product and with the usage of non-separable states to establish the complex interdependent relationships between the varied representations. They also allow for handling different forms of user interaction and apply advanced principles of Polyrepresentation [63], which states that a document is more likely to be relevant if it's relevant to cognitively and functionally different representations. Multiple representations of documents and queries such as user reviews or reformulated queries can potentially exhibit user dynamics. The user dynamics can provide relevant and non-relevant facets of searchers' information needs. The idea of polyrepresentation in combination with the quantum framework can model user dynamics in Hilbert space.

Earlier work in [211] based on objectives referred from Quantum Theory (QT) has manifested methods of standardising aspects of IR focussing on an extensive theoretical basis in which a search activity can be thoroughly defined and well organised, and a scientific foundation inspired by the operational procedure in QT. It was eventually underlaid that there is a possibility for QT methods which acts as a major role in clarifying the above IR problems (of interpretation and paucity of scientific technique) manifested. Motivated by these distinctive relationships and on undertaking to interpolate these methods such as Information Foraging theory to QT for interactive IR. It was established that the search solicits to be re-examined from a viewpoint quite diverse from how it is traditionally distinguished [11] in order to infer the utility, feasibility, and procedure of integrating IFT models. Considering search in this new manner also evinces approaches for re-defining the concept of search [12]. The main goal of this research can be equalised to being able to properly refer to IR as *user-oriented IR* by underlying a distinct formulation of search and inferring scientific methods for the investigation of user information needs, so it can be in all reverences, an interactive search.

2.3.7 Realm of IFT

A forager while searching¹¹ in a patchy environment [40, 164] encounters two types of fallacy: equivocation pertaining to the idiosyncrasy of the current patch and risk allied with the background opportunities. We infer that the order in which the forager deals with these fallacies influences the decision of whether to remain at the current patch. The order effect is formalised with the non-Kolmogorovian model and quantum probabilistic framework [33]. We show this by applying the mathematical formalism of quantum theory including some features of *non-Kolmogorovian probability theory* to the Information Foraging task, in particular, to forager's decision in information patch.

¹¹'Searching' context is finding information objects such as text, image, etc. in Web search

Earlier studies [13] have shown the violation of the total probability law in the well-known double-slit experiment. We can advance this interpretation by discussing the violation of the total probability law in the domain of Information Foraging Theory. Based on [236], we hypothesise that when a forager looks into the information patch and asks whether the given information such as ‘IFT’¹² exists ($I = 1$) or not ($I = 0$). They collected statistical data and determined the probabilities $P(I = 0)$ and $P(I = 1)$. In another demonstration¹³, we then provide information goal (G)¹⁴ to the (other) foragers before knowing them which information patch to look for. Using conditional probabilities for comparison with the quantum approach, we evaluate the probabilities $P(G = 1)$, $P(G = 0)$, $P(I = 1|G = 1)$ and $P(I = 1|G = 0)$. The probability that the information patch contains information ‘IFT’ is determined as

$$P(I = 1|G = 1)P(G = 1) + P(I = 1|G = 0)P(G = 0). \quad (2.12)$$

A simple application of classical probability theory implicitly evaluates that it is equal to $P(I = 1)$. Although, one can interpret the value by Eq. 2.13 is smaller than $P(I = 1)$;

$$P(I = 1) \neq P(I = 1|G = 1)P(G = 1) + P(I = 1|G = 0)P(G = 0). \quad (2.13)$$

The probability of $P(I = 1)$ in Eq. 2.13 is estimated in the context of ‘foragers did not know their information goal before knowing the information patch’. Contexts ‘a forager knows their information goal’ and ‘a forager did not know their information goal’ is different. Let’s consider first context is represented by S_G and the latter by S_{-G} . Substitute these two contexts by $P_{S_{-G}}(I = 1)$ and $P_{S_G}(I = 1)$ in both sides of the above equation. However, the solution of $P_{S_{-G}}(I = 1) \neq P_{S_G}(I = 1)$ appears to be now natural. To describe this solution mathematically, we need a complete probability space that describes the two contexts S_G and S_{-G} . To find such a typical probability space needs extensive knowledge about the physical and chemical structure of a forager (human being) will be needed, so it is very rigorous to find such a typical space, as in the circumstance of the hidden variables theory. Here, the joint probability distribution is, in general, not defined. Instead, we define quantum conditional probability as the probability with respect to the state achieved as the revision of the initial

¹²I symbolises for IFT, where ‘IFT’ patch exists as $I=1$ otherwise $I=0$

¹³The examples here adopted and modified from [236]

¹⁴Information goal refers to the users’ desired patch, where $G = 0$ is the user unable to reach his/her information goal, otherwise $G=1$

state ψ after the first measurement (and essentially dependent on the first measurement outcome). The quantum states of contexts¹⁵ S_G and S_{-G} are created as ρ_{S_G} and $\rho_{S_{-G}}$.

$$P(I = 1|G = 0) = \frac{\|P_1 P_0 \psi\|^2}{\|P_0 \psi\|^2}, P(G = 0|I = 1) = \frac{\|P_0 P_1 \psi\|^2}{\|P_1 \psi\|^2}.$$

The order effect takes place iff $\|P_1 P_0 \psi\|^2 \neq \|P_0 P_1 \psi\|^2$, or $\langle [P_1, P_0] \psi | \psi \rangle \neq 0$.

By the conventional description of quantum mechanics, the corresponding probabilistic structure cannot be depicted classically in a Kolmogorov space. This illustration aims to bridge the gap by addressing the users' cognitive factors during search (or foraging) tasks by combining the quantum theory with its mathematical formalism and Information Foraging theory. However, we assert that the affinity between classical and probabilistic interpretations is still passionately debated.

2.4 Conclusion

We embark on a comprehensive insight into information retrieval modelling and its generalisation using the quantum probabilistic framework. The passage of user aspects in IR, in particular, certain traditional and new methods are also surveyed, i.e., Ostensive IR and IFT. We also present some recent methods that employ IFT for recommender systems, search, and a few IR tasks. Also, we highlight the advantage of incorporating IFT constructs in quantum-inspired IR models in order to support user aspects.

The notion of quantum-inspired IR methods set forth a generalisation lookup way for classical models of information retrieval and introduced its extension via the usage of quantum probabilities. This extensive approach uses quantum probabilities to explicate user interaction across the searcher queries (information needs) and documents (or images), or even how the user information needs evolve or shape under a specific situation. The most applied modelling method follows the Quantum language model. However, there are certain approaches such as polyrepresentation which are yet to be employed on any benchmark datasets. The representation framework for the user information need based on the Hilbert space formalism stands as a standard tool to integrate with a generalised model. In the subsequent chapters of this dissertation, we will cover our contributions to new quantum-inspired IR models and approaches for specific tasks in IR.

¹⁵The value of ρ_{S_G} and $\rho_{S_{-G}}$ can be computed from [113]

Chapter 3

Using Quantum-inspired Word Embedding for Modelling Searcher Behaviour

3.1 Generalising Word Embedding using Quantum Probabilistic Framework

Word embedding has been extensively investigated [23, 47] to form the basis of a learned representation and understanding of the text in a vector space. The underlying notion of word embeddings is to employ the co-occurrence of words as semantic correspondence among words. Word embeddings are also used as text (or word) features [208]. It has been used to enhance the language models [69], which is the applicability of word embeddings. However, amplifying (or simplifying) the mechanism of word embeddings could lead to better results in comparison to traditional word embeddings. It is essential for word embeddings to capture the meaning of combined words such as the sentence ‘strong coffee’. And, the high frequency of locus of the words ‘strong’ and ‘coffee’ lapse to indicate that they are negatively correlated.

In the information retrieval community, formal models using the mathematical formalism of quantum theory have been exercised to tackle cognitive aspects among words [201, 241], which follows the first and foremost foundation of quantum theory for information retrieval described in [211]. We employ cognitive aspects to word embedding based on complex-valued embedding network [124]. It assumes that users do not liken a unique polarity or sentiment to each word, whereas a term embraces the global polarity of combined words based on the other entities it is paired with. This follows the fact from quantum theory that the action of tiny particles which dwell in every possible state simultaneously hinders each

other and gives rise to new states based on the relative aspects, and is analogous to words that occur in similar contexts and tend to have similar meanings.

In this chapter, we describe a generalised approach for complex-valued word embedding – a lightweight method for training such an embedding network. Our investigation into quantum-inspired word embedding set forth an algorithmic approach that optimises a deep neural network [206] whilst dealing with pre-trained embedding models in a way that makes the training steps computationally efficient.

3.1.1 Related Work

The very inception of Word2Vec for generating word embeddings which uses a two-layers neural network to process text in vectorised form. There are a set of algorithms such as a continuous bag of words and skip-gram that capture the co-occurrence of words sequentially instead of delineating all co-occurrence counts at once like singular value decomposition. The word embedding mechanism places words in a low-dimensional space in which it approximates the distance among words for semantic similarity, and linear relationships among the word vectors. Recent advancements in word embedding approaches cater to actual morphological language structures [142, 160] that act as an enhancement to linguistic tasks. Additionally, an unsupervised learning technique for word representation has been introduced in [160], so-called global vectors of word representation (GloVe), which is trained on aggregated global word-level co-occurrence stats from a large corpus of texts. The output word representation was found to be effective on a diverse set of NLP and IR tasks [183, 182, 51], where word vectors exist as linear substructures in the vector space. Word aspect can be altered to diverse character-level n-grams that compose it, such as FastText [29], a character-based embedding model to cater to the language characteristics or rare words that are unseen in the training corpus. FastText is a fine-grained embedding model that amplifies vocabulary analysis using character n-grams.

Representation of combined words using embedding models is more driven toward relationship among words. However, [201] tackles the dependencies between words (or text segments) and modelled it as a quantum mixed state, represented by a density matrix where off-diagonal elements in it delineate word relationships in a quantum description. Traditional word embeddings allocate in a vector space (or embedding space), whereas quantum-inspired word embedding simplifies the feature space to operate in real-valued Hilbert space. Word or text segments are in the representation space of real-valued vectors (in the form of matrices), it is due to the paucity of appropriate linguistic features contributing to the imaginary part. Manifestation of quantum events cannot be simply delineated without complex numbers [110, 118], as these models are theoretically confined. From the perspective

of information retrieval, document collections are a general aspect of a retrieval task, where a theoretical framework underlying quantum theory so-called QWeb [3] models document collections. This framework assumes a term as a substitute for ‘whole meaning’ which is regarded as a concept delineated as a state in a Hilbert space and the concept mixture is described as a superposition of the concept states. All of the concepts with their corresponding complex phases are associated with the scope of interference among concepts. However, QWeb is yet to be applicable to any NLP or IR tasks. The applicability of quantum theory for IR tasks in [170] employs term vectors and pairs of it using mixtures and superposition to delineate information needs in a real-valued Hilbert space. This meaningful description of queries and documents performs like the standard Okapi BM25 ranking algorithm.

In our case, we employ a quantum-inspired approach to word embeddings [124], where words are a linear series of latent concepts provided words rendered with complex weights and combined words as a complex series of word states (words in superposition states). The terms in a series of words can be delineated either in a mixed or superposition state which adheres with [3] besides their conjecture of terms are characterised as ‘entities of meaning’ in QWeb.

3.1.2 Method

We conduct an extensive analysis of the complex-valued word embedding described in [124] provided our methodical approach follows sentence-based interpretation. We assume a sentence as a blend of words given a word is regarded as a quantum state and is of two folds:

- Co-occurrence of words having real-valued amplitudes captures the low-level information.
- A word paired with another word can give rise to varied contexts associated with it and complex phases of combined word delineate the emergent meaning.

Hence, the meaning of combined words prevails in the interference between words and can be captured implicitly in the representation of the density matrix.

The technique introduced in [124] for classical mixture of sentences, where each word vector is depicted by \vec{w} as $\vec{w}^T \vec{w}$ (similar to Dirac’s nomenclature $|w\rangle$ and $|w\rangle \langle w|$) magnifies the computational capability to train such a deep neural network parameterised via quantum probability constructs. We employ the established pre-trained word embedding models [29, 143, 142] as prerequisite word-level features for the experimental settings. However, these embedding models consist of 300-units dimensions per term, or alternatively a 300×300 complex embedding matrices. Such a higher dimensional quantum structured

objects recline toward smaller batch sizes and monotonous increment in random access memory (RAM), and hence the model requires a longer time for training due to the input word vector $|w\rangle$ of dimension $(||w\rangle|, ||w\rangle|)$ instead of $(1, ||w\rangle|)$. Having the input word vectors of high dimensionality, we reduce the embedding dimension using approaches described in [148, 173] by reproducing both methods in a combined manner. [148] approach is to deduce the mean energy of the vectors vested in the model, which increases the discrimination among vector indexes. Our method performs reduction using principal component analysis (PCA) [238, 233] to detect the top components which contribute to the major variance ratio among each unit of the word vectors of the entire model vocabulary. Then, it filters those identified top components and applies additional PCA transformation (to halve the size of the dimensions) on the pre-trained embedding models. After reducing the size, the original model is outperformed on optimising the training process for the quantum-parameterised neural network.

In order to leverage the PCA technique [233], a small number of new features, which are linear combinations of the old features, elucidate most of the variance in the data. The principal component directions are the directions in the feature space in which the data is the most variable. The PCA process follows the below steps:

1. To standardise the range of continuous initial variables to ensure that each variable is given equal weight in subsequent analyses.
2. To determine correlations by estimating the covariance matrix.
3. To determine the principal components by estimating the eigenvectors and eigenvalues of the covariance matrix.
4. To ensure which principal components to retain, a feature vector can be created based on the eigenvalues, which represent the amount of variance explained by each principal component.
5. The data can be recast along the principal components axes.

Mathematical Interpretation

Let the m features in our input embedding model be $x = (x_1, x_2, \dots, x_p)$. We define a new feature, the first principal component score, by

$$p_1 = \phi_{1,1}x_1 + \phi_{2,1}x_2 + \dots + \phi_{p,1}x_p = \phi_1^t x$$

Here, the coefficients $\phi_{j,1}$ are the *loadings* of the j -th feature on the first principal component. The vector $\phi_1 = (\phi_{1,1}, \phi_{2,1}, \dots, \phi_{p,1})$ is called the *loadings vector* for the first principal component.

We want to find the loadings so that p_1 has maximal sample variance. Let X be the $n \times m$ matrix where $X_{i,j}$ is the j -th feature for item i in the dataset. X is just the collection of the data in a matrix. Assume each of the variables has been normalised to have a mean zero, i.e. the columns of X should have zero mean.

A short calculation shows that the sample variance of p_1 is then given by

$$\text{Var}(p_1) = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m \phi_{j,1} X_{i,j} \right)^2.$$

The variance can be arbitrarily large if the $\phi_{j,1}$ are allowed to be arbitrarily large. We constrain the $\phi_{j,1}$ to satisfy $\sum_{j=1}^m \phi_{j,1}^2 = 1$. In vector notation, this can be written $\|\phi_1\| = 1$.

Putting this together, the first principal component is defined by $p_1 = \phi_1^t x$, where ϕ_1 is the solution to the optimisation problem

$$\begin{aligned} \max_{\phi_1} \quad & \text{Var}(p_1) \\ \text{subject to} \quad & \|\phi_1\|^2 = 1. \end{aligned}$$

Thus, ϕ_1 is the eigenvector corresponding to the largest eigenvalue of the covariance matrix $X^t X$.

We similarly define the second principal direction to be the linear combination of the features, $p_2 = \phi_2^t x$ with the largest variance, subject to the additional constraint that p_2 be uncorrelated with p_1 . This is equivalent to $\phi_1^t \phi_2 = 0$. This corresponds to taking ϕ_2 to be the eigenvector corresponding to the second largest eigenvalue of $X^t X$. Higher principal directions are defined analogously.

Alternatively, the techniques introduced in [148] can be enhanced using the approach described in [173], we also extended the latter approach using our proposed dynamic components finder, which is to identify the number of components to be removed from the model given a threshold. We empirically analyse this case by removing the principal components which contribute to 20% of the variance and delineate either 6 or 7 components based on which the embedding model is being used. It can depict meaningful results with the model transformation approaches and is independent of the algorithm upon which an embedding model has been trained. We report the two steps of our algorithmic operation, where Algorithm 1 dynamically generates the γ factor as a counter to detect the number of

components needed to be removed, and Algorithm 2 shows the main transformation process based on [173].

Algorithm 1 Dynamic Model Discrimination

Input: V ▷ Embedding model
Output: V ▷ Compressed embedding model with γ as dynamic factor

- 1: **procedure** DISCRIMINATE-MODEL(V , threshold = 0.2)
- 2: $\mu = \text{MEAN}(V)$
- 3: **for** $n = 1, \dots, |V|$ **do**
- 4: $\hat{V}[n] = V[n] - \mu$
- 5: **end for**
- 6: $p_{1\dots d} = \text{PCA}(V)$ ▷ p_i , where $i \in 1, \dots, d$ refers to identified principal components
- 7: $\gamma = 0$
- 8: partial_ratio = 0.0
- 9: **for** p in p_i **do**
- 10: $\gamma += 1$ ▷ variance refers to the eigenvalues of the covariance matrix
- 11: **if** partial_ratio + $p.\text{variance_ratio} \geq \text{threshold}$ **then**
- 12: **break**
- 13: **end if**
- 14: partial_ratio += $p.\text{variance_ratio}$ ▷ variance_ratio is the percentage of variance
- 15: **end for**
- 16: **for** $n = 1, \dots, |V|$ **do**
- 17: $V[n] = \hat{V}[n] - \sum_{i=1}^{\gamma} (p_i^T V[n]) p_i$
- 18: **end for**
- 19: **return** V
- 20: **end procedure**

where V = the model being transformed, depicted by a pair of key-value of the term with its corresponding vector representation, variance ratio to normalise the principal components for the embedding vectors to be uniformly distributed in all dimensions, threshold = threshold for the major variance ratio interval.

Algorithm 2 Model Reduction

```

1: procedure REDUCE-MODEL( $V$ )
2:    $n = \frac{1}{2} \times (|V|)$ 
3:    $V = \text{DISCRIMINATE-MODEL}(V)$  ▷ From Algorithm 1
4:    $V = \text{PCA-TRANSFORM}(V, n)$ 
5:    $V = \text{DISCRIMINATE-MODEL}(V)$ 
6:   return  $V$ 
7: end procedure

```

where V = the model being transformed, depicted by a pair of key-value of the term with its corresponding vector representation.

3.1.3 Experiments

We detail the evaluation part of our proposed algorithmic approach that applies to the complex-valued word embedding [124]. The aim is to investigate the performance when a classifier combines with varied words (or character) embedding models. We demonstrate our evaluation effectiveness on benchmark datasets of question classification as one of the textual tasks.

Dataset

We consider the question classification task for the proposed algorithmic approach to evaluate and selected the two benchmark datasets for the sake of fair comparison with models manifested in [124], namely TREC-10 [31] and SST-2 [198] datasets shown in Table 3.2. Another set of datasets consists of pre-trained word embedding models (which are actually a feature) - GloVe [160] and FastText [29]. However, our interpretation here is to consider these pre-trained word embeddings as datasets reported in Table 3.1 as it supports word-/character-level embeddings, and we also inline these datasets separately to validate our findings.

The notion of using GloVe and FastText embeddings is of the fact that the first operates alike word2vec embedding, where each word acts as one distinctive representation of a vector. In FastText embedding, it is made up of varied character n-grams that remit the constraints whilst dealing with unknown and rare words during model training, and this is in an eventual multi-linguistic scenario, as it is agnostic to conditions on morphological connections.

We apply our algorithmic approach described in Algorithm 1 and Algorithm 2 on embedding models as an input reported in Table 3.1 to complex-valued word embedding

Word Embedding Models	Vocabulary Size	Corpus Size
(A) GloVe-300d [160] †	400 thousand	6 Billion
(B) FastText-300d ‡	2 Million	600 Billion
(C) GloVe Common Crawl-300d †	2.2 Million	840 Billion

Table 3.1 List of the pre-trained word embedding models where †depict GloVe embeddings and ‡depicts Fasttext embeddings. The word embeddings vector is of 300 dimension.

Datasets	# Records	# Classes
TREC 10 Question Classification	5,952	6
SST-2	70,042	2

Table 3.2 Overview of the Datasets

network [124], which after transformation results in five-word embedding models i.e., for each embedding model (B) and (C) with vector length of 150 (150d) and original embedding of vector length 300. So, the original embedding model and transformed one for each of (B) and (C) lead us to four embedding models and one for (A).

3.1.4 Results

The proposed algorithmic approach applies to the complex-valued word embedding model which inputs the word or character-level embeddings and this evaluation uses the question classification datasets. In comparison performance of the varied embedding models in Table 3.3, we represent the words depicted as real-valued numbers by ‘R’ and the reproduced mixture model (based on [124] by ‘M’. Also, our analysis reflects that FastText (B) embedding model is on par in comparison to other models, in particular traditional word embeddings. However, the performance of GloVe embedding models is even higher when used as an input to the complex-valued word embedding model. We also compare it with the reduced versions of the embedding models on how closely these models are with the original pre-trained embeddings (i.e., vector length of 300). This led to an interesting finding on a pre-trained embedding model when reduced to a vector length of 150, which requires less computing resource and time for training and still achieves similar performance to the original embedding models i.e., lower complexity among 150 and 300 units of pre-trained versions. Despite the reduced version (50%) of embedding model (C) which was trained on a large corpus containing 840 billion tokens, it does not deteriorate the entire performance in comparison to its original pre-trained embedding model. Thus, the explored analysis invokes

Word Embeddings	Reconstructed Dataset	Classification Accuracy			
		SST-2-R	TREC-R	SST-2-M	TREC-M
(A)		78.47 %	79.80 %	82.14 %	75.48 %
(B)		79.29 %	79.30 %	81.73 %	84.84 %
(B) - reduced		78.86 %	82.30 %	81.83 %	85.20 %
(C)		79.10 %	80.50 %	82.42 %	84.20 %
(C) - reduced		78.42 %	80.00 %	82.21 %	85.48 %

Table 3.3 Evaluation Results across different word embedding models. ‘R’ depicts the updated word embedding model with word representation as real-valued numbers. And, ‘M’ depicts the reproduced embedding mixture based on [124]

Word Embeddings	Reconstructed Dataset	Time-to-Train (in secs)			
		SST-2-R	TREC-R	SST-2-M	TREC-M
(A)		31s	8s	974s	51s
(B)		31s	8s	974s	51s
(B) - reduced		20s	5s	332s	18s
(C)		31s	8s	972s	51s
(C) - reduced		20s	5s	332s	18s

Table 3.4 Training time per epoch for each word-embedding model across varied text classification datasets listed in Table 3.2

the trade-off between complexity and embedding size against performance and less reduction in accuracy.

Also, the time taken during training steps for each of the pre-trained embeddings and their corresponding reduced version are reported on a scale of ‘time per epoch’, and is shown in second for each model per epoch in Table 3.4.

We created a publicly available gist that contains execution logs of training time steps for these above four datasets which are SST-2-R¹, TREC-R², SST-2-M³, and TREC-M⁴.

¹<https://bit.ly/2Oq1WR1>

²<https://bit.ly/2Ae72wx>

³<https://bit.ly/2LOw0qG>

⁴<https://bit.ly/2LrFQzs>

3.1.5 Conclusion

Connecting classic word embeddings with quantum-inspired information retrieval framework [211, 171] constructs allows us to understand the emergent meaning of combined words. The proposed algorithmic technique has explored an approach on how to reduce the size of embedding models that optimises the overall training of quantum-parameterised word embedding model and is more effective than the state-of-the-art complex-valued word embedding and baseline models validated on the question classification and sentiment analysis datasets. The presented method employs large pre-trained word embeddings as part of the experimental pipeline and a strong comparison of it reflects a demonstrable enhancement for generalised word embedding models.

Though our approach behind generalising word embedding models merely stems from a critical problem on word-level semantics [108] which superposes two varied word states. Our findings reveal insufficient support to reject our notion of generalising the complex-valued word embedding model, however, on-par performance in terms of classification accuracy and training time. We foresee an aspect of each word in a sentence i.e., the word weight can leverage the performance of learning tasks and so the accuracy of a bigger model. Further analysis of the introduced γ , a dynamic real-valued factor remains part of future work. This factor can be investigated in rating the inductive term importance within sentences depending upon whether γ increases or decreases. This dynamic rating can reflect the context of a term present in a sentence.

3.2 Quantum-inspired Reinforcement Learning-driven Information Seeker

Searchers across the webpage follow or consume information successively traced via hyperlinks or cues, where the consumed information may be irrelevant if the generalised information (or information diet) lies ahead in the search trail. Such an information environment possesses uncertainty [165, 44]. Usually, searchers keep on exploring information whilst interacting with the system and emanation to information goal (optimal information diet) inhibited by their involvement in an information seeking process. Thus, there requires a mechanism that can guide the foragers during their search process to manifest their information appetite (optimised information diet). The key process of finding information in web search is to involve in user interaction that can improve the search performance, and the information foragers' search experience and satisfaction [207, 30, 129]. User-level aspect such as action during search reflect their altered behaviour and belief states due to severity

across the information environment [228]. Recent work introduced a behavioural approach using reinforcement learning (RL) [39] to represent the action behaviour of users, where they extrapolate the actions to a certain state via a policy in two constructs - representation of user action and its corresponding transformation. In our work, we believe that the searchers (or information foragers / users) [237] cognitive ability can be enriched via guidance in finding the information, and so we consider the information forager as an RL agent based upon Information Foraging Theory (IFT) [166]. This helps us understand how well a searcher can learn during the process of locating information. Moreover, the learning ability of a searcher can be informed by the reinforcement learning approach by allowing an independent selection of search space in an uncertain environment. RL approach can be enhanced using the notion of IFT for information seekers (or foragers) in order to improve the trade-off between exploration via perpetual steps in the information space, and exploitation via the encountered information patches. The latter trade-off component i.e., exploitation happens when the searcher finds a particular information patch to be of interest due to its strong information scent. We reckon that such a trade-off can be a key factor for understanding searchers' belief [228] with uncertainty i.e., risk and ambiguity during the search process [237]. Thus, this behavioural framework of Information Foraging Theory is primarily sequential. So, the attenuation of risk and ambiguity constructs cannot take place simultaneously, which directs to an underlying extent how good the exploration-exploitation trade-off can be for the searcher. This magnifies the Information Foraging perspective of information seeking to meet with the growing field of quantum theory [237].

Searcher with access to a large amount of information on the Web in the search process leads them to acquire new skills and knowledge. However, searchers' information needs are open-ended, usually diverse, and seldom unclear at the very start. Such types of information needs commonly involve multiple queries and rich interactions. Therefore, we aim to model the users' information need that alters during a search based on the Information Foraging process.

Users' information needs are instinctively complex and intense evolving nature, the Information Foraging process is insistent on both the searcher and search system. Thus, our center of attention is to incorporate semantic information in modelling the information forager using the quantum probabilistic framework. In this sub-chapter, we propose a reinforcement learning approach parameterised using quantum probabilities constructs are of two folds:

- The foragers' behaviour is modelled as a reinforcement learning task provided action-selection (or policy) is described using an Actor-critic mechanism, that amplifies the agent's experience in a text query matching.

- The reinforcement learning approach learns the policy in which the query representation is parameterised using quantum language models, and incorporates words with multiple meanings.

3.2.1 Related Work

Information foragers during a search are mostly involved in a decision making process [55], for instance, one of their actions leads to an adverse consequence (or negative payoff) and so this user action will be discarded in the future. However, if foragers' action leads to a positive outcome (or reward) then it will occur again. This scenario of decision making renders similar features of reinforcement learning. We employ RL modelling approach which is apparent to prevail in searchers for learning information whilst searching. White et al. [229] investigated modelling the search context in a typical search system that improves search intention and manifests the sophisticated query reformulation. Also, reinforcement learning has a multitude of techniques that can allow the system to follow the search context and relevance feedback at the same time. This even further provides the search system to deal with exploration (expanding the search between varied topics) and exploitation (moving intensively into particular sub-topics) which has been supportive in information retrieval [247, 192]. Generally, exploration and exploitation mechanisms are profoundly used in tasks allied with information retrieval and recommender system tasks, such as duelling bandits in IR [244], foraging approaches for learning the entire duration of search [58] and in text recommendation [20]. However, the applicability of reinforcement learning in information retrieval systems [255] records users' interests and skills, whilst in a generic search task, the searchers in a given search session rely upon the entire enhancement of the search results instead of participating in an arbitrary search session.

The exploration-exploitation trade-off in reinforcement learning can be extricated using Information Foraging Theory due to its sequential nature. Information Foraging Theory is a behavioural framework for information retrieval. Our work focus on an information seeking task, where searcher behaviour can be characterised using IFT strategies in RL. It has been applied to describe the information needs of searchers and their actions [42] using the information scent model. Also, the evaluation of the information scent-based model was introduced in [43] to examine and predict the usefulness of a website through the website's scent. Liu et al. [129] evaluated a user classification model for content-based image retrieval based on IFT to characterise the users' preferences and behaviours during the search. Additionally, the evaluated model covers a broad range of user interaction features collected from the screen capture of varied users whilst searching. IFT constructs are key components in determining searchers' behaviours. It is not only limited to textual tasks in

IR but also in image recommendation [96], which uses foraging strategies to recommend image items through inspection of visual patches and characterises how searchers follow recommended items. A similar scenario of user-item interaction has been introduced in a query auto-completion task [99]. Through this, having a reinforcement learning model based on IFT can enhance the explainability of searchers' behaviour [99, 152].

Dynamic searchers' actions under a certain state using the reinforcement learning approach can be modelled, however, it does not encapsulate all of it under one framework. To fulfil the expressiveness in a single framework, we employ the quantum probabilistic framework of information retrieval [211, 171, 170]. The conjunction of quantum theory and information retrieval is to understand and characterise the interaction between a searcher (the observer representation with quantum constructs) and the information object under observation [211]. Also, probabilistic features are common among quantum theory and information retrieval, that is an observation of agreement for the preface of conditional probabilities connected with interference effects dominating to few contextual measures (subjective or cognitive character) when consolidating different information objects⁵. We employ constructs of quantum probability to delineate information needs (or queries) with eigenvectors (or state vectors), and the corresponding eigenvalues and the probability of finding single eigenvalues or information objects as a measure of the degree of relevance to a query [170]. A very interesting scenario of the reinforcement learning framework parameterised using quantum theory can [60] introduced an algorithmic strategy to deduce favourable user actions in a dynamic environment.

We formulate the information seeking as a reinforcement learning task driven by Information Foraging followed by encapsulation under the quantum probabilistic framework. This led to a quantum-inspired reinforcement learning (qRL) framework for the interplay between the information forager (or seeker) and the search space (or search system).

3.2.2 Information Seeker as Reinforcement Learning Agent - Hypothesis

The hypothesis for the information seeker is that during the search process, the searcher (or forager) deals with multiple actions prior to selecting any of them, with an undetermined reward. Foragers continually explore each search result thereabout to estimate the optimal information patch depending upon the reward. This recalls the interpretation similar to reinforcement learning where the search task or search session, involves an agent interacting with the information environment, and the searcher (or agent) observation is cost-driven.

⁵<https://www.newscientist.com/article/mg21128285-900-quantum-minds-why-we-think-like-quarks/>

Wittek et al. [237] found uncertainty in searchers' decision making for an information seeking task, where information seeker act as a forager. This leads us to sequential decision making which harmonises with Information Foraging Theory, and so we refer to it as *Reinforced Foraging*. The RL agent in an uncertain information environment assessing actions that are positively rewarded (costs incurred by searchers whilst their judgement among actions) can likely renovate the foragers' choice in finding the information. From the viewpoint of Information Foraging Theory, the resemblance of positively rewarded actions in RL can be regarded as exploitation in IFT whereas the readily available actions as exploration given the information must be expanded within a patchy environment. Reinforcement learning inscribes the primordial aspect to 'learn by doing with delayed reward' which is coincident with information seeking (in particular, user interaction in IR and recommender system tasks), delineating an intuitive interpretation of the foraging session of a searcher. Foragers' objective is to rapidly find the relevant information patch (document, image, etc.). However, the information seeker is mostly unaware of the rewards from trailed information patches and they continue to explore each one of them. Foragers during exploration on interaction with the search system locate the relevant information, which delineates the reward distribution (information scent patterns) among information patches. Usually, the accessed information patches with minimum reward can evince an optimal patch that the forager has spent less time upon whilst exploration. An information seeker during exploration spending less time on inspection of information patches results in partially relevant information, and the foraging process delineating the rewards distribution among patches leads to exploitation, which has less (moderate time spent for exploitation and it may not be quite relevant information patch) than the optimal rewards. Therefore, the longer a seeker/forager explores, the chances of obtaining near-accurate information of all information patches become stronger. Such demystification of the information seeking process from the viewpoint of RL and IFT paves the approach to model user foraging behaviour in which their cause could be confusion, uncertainty in decision making, and information overload.

3.2.3 The Framework

The proposed quantum-inspired reinforcement learning approach models the forager's action in an information seeking scenario, where the task is to match a query for a given document in which the forager's actions are expressed in queries (information needs). The notion of an RL agent is to maximise their reward by finding an optimal foraging strategy. Also, the forager's in a given environment have a definite set of optional information sources, and they have the preference to affix an incisive type of information patch into their diet. The distribution of incisive information patches may include information that the forager could

plausibly extrude, due to counterfactual states in deciding between which information patch (for instance, document d_1 and d_2) comprises essential information. In this framework, we hypothesise that the information environment is uncertain with dynamic parameters during a forager's search trail. This situation for the forager makes it stiff to extricate patches and so exploit his/her experience to learn the environment. The prolonged learning happens to be complex at the cognitive and dynamic level provided the forager's quest is to find the most relevant documents.

We employ the Actor-critic policy gradient technique [132] to model foragers' dynamics due to their sequential behaviours that procreate a continuous state representation. The mathematical description of a forager's action (or state) can be represented using the quantum superposition state under the updated state vectors, based on the envisaged interaction, it can be attained via random observation of the imitative quantum state which follows the collapse principle of quantum measurement [211]. The probability of an action state vector is the probability amplitude updated in parallel and can be estimated based on reward. This emerges as an intrinsic aspect in traditional reinforcement learning algorithms which are representation, action (in parallel), policy, and operation update. The quantum measurement of a forager decision state in selecting a document (action) whilst an information seeking task is uncertain and ambiguous [237]. To accomplish the measurement of an observable delineating probabilistic action can be represented by an operator \hat{O} , with two state vectors consisting of $|0\rangle$ and $|1\rangle$. The measurement of a quantum system for an observable (\hat{O}) in a corresponding superposed quantum state $|\psi\rangle$, is a measurement in the superposition state. During the measurement in quantum state $|\psi\rangle$, it would collapse into one of its basis eigenvectors (or state vectors) $|0\rangle$ or $|1\rangle$. The collapse of a quantum state in either of these basis states does not change the prior and so it can not be obtained with certainty. This quantum system can only provide information that $|0\rangle$ corresponds to the measurement with probability $|\alpha|^2$, and the probability $|\beta|^2$ to measure $|1\rangle$, where α and β depicts the probability amplitudes.

The proposed quantum-inspired reinforcement learning (qRL) framework for information seeking based on Information Foraging theory applies under dynamic search sessions. The Actor-critic [132] method describes the reinforcement learning agent to conjunctly encodes the action and state spaces, and the information environment conferring documents. The constructs of an RL agent for the Actor-critic method are parameterised using quantum probabilistic framework [211] and quantum language model [201].

This qRL framework investigates a typical matching task, specifically, semantic query matching where a set of candidate queries will be matched to the document given in the sample set in semantic Hilbert space (SHS) [216]. SHS is a complex vector space of words,

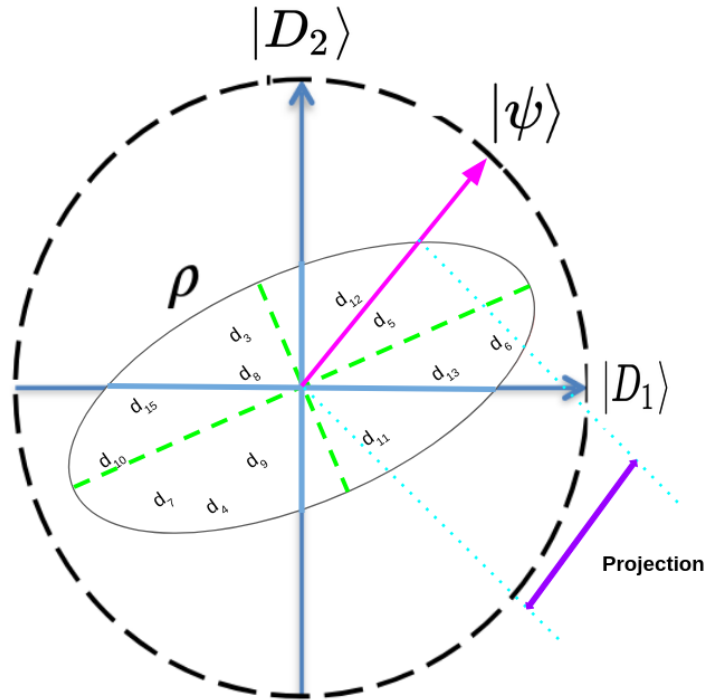


Fig. 3.1 Selection of Documents in Hilbert Space

where multiple words together lead to a linear or non-linear creation of amplitudes and phases, characterising combined words at a different level of semantics. A word w_i in SHS can be represented by a base ket vector $|w_i\rangle$. The representation of combined words is a superposition of word vectors, which are encoded in the probability amplitudes of the commensurable base vectors.

Preliminaries

To formulate the query matching task for the qRL framework, the first and foremost step is to prepare the constructs of reinforcement learning. A typical reinforcement learning follows Markov decision process (MDP) and formed on a finite state, of which their constructs comprises of five components - the state at a certain time t (s_t), possesses an action at time t (a_t), the transition probability ($p_{ij}(a_t)$) i.e., one state s_t to another state (s_{t+1} through action a_t for all $t \in (i, j)$, a reward function r in which $r : \Gamma \rightarrow \mathbb{R}$ provided $\Gamma = \{(i, a) \mid i \in s_t, a \in a_t\}$ and an objection function C . We follow a range of notations as listed in Table 1.2. The redux of the proposed framework adheres to Hilbert space formalism, where Hilbert space is denoted by \mathcal{H} . Also, we mainly employ and borrow mathematical interpretations of Tensor space (such as tensor product) based on the quantum-inspired language models [251, 249]. In an n -dimensional Hilbert space on a real plane denoted by \mathbb{R}^n , where the base

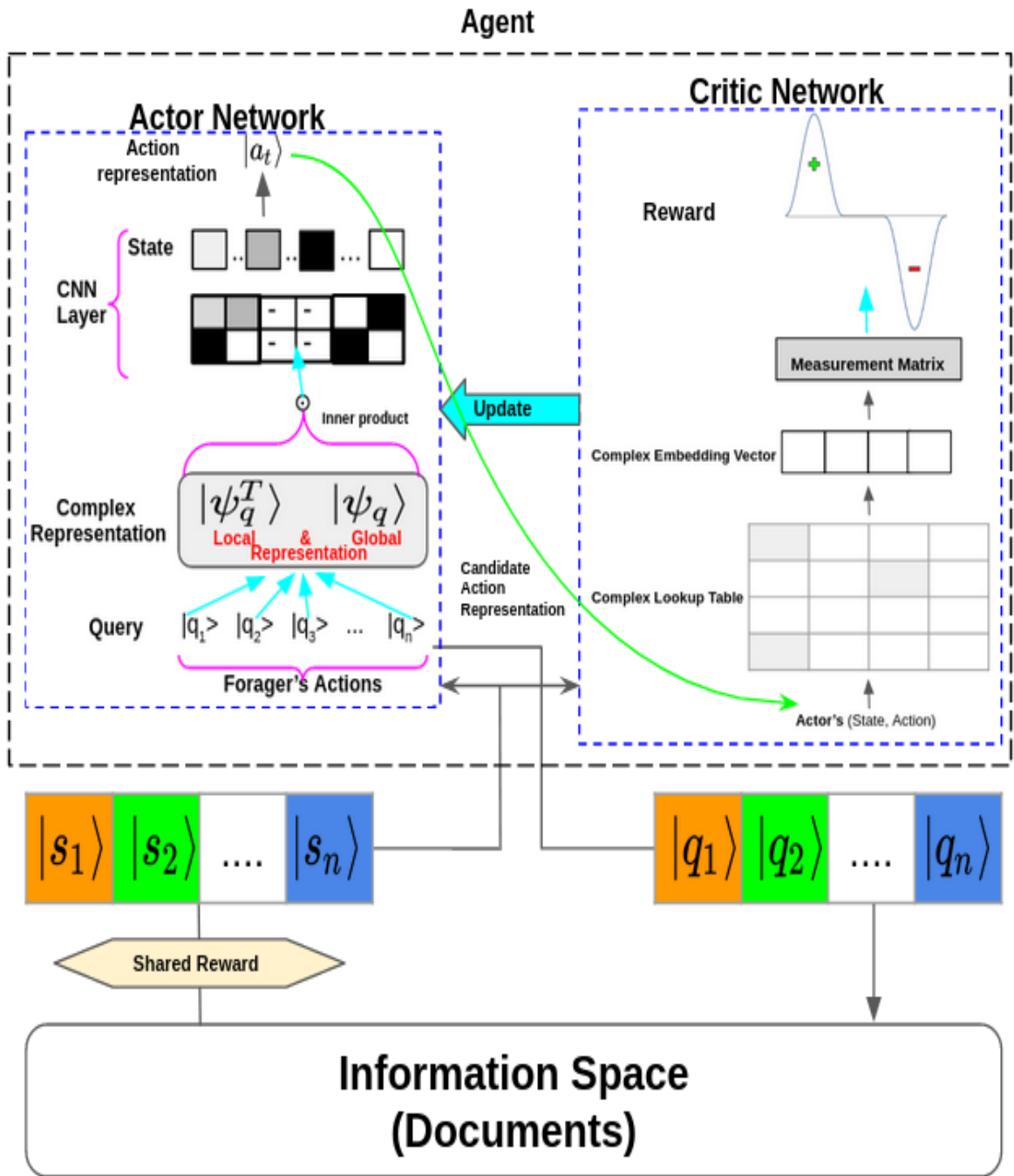


Fig. 3.2 Architecture of Quantum-inspired reinforcement learning framework

vectors represented by $|\phi_i\rangle_{1,2,\dots,n}$ are term vectors (or word embeddings). A word vector $|w\rangle$

represents a linear combination of the base vectors as a coefficient, where

$$|w\rangle = \underbrace{\sum_{i=1,2,\dots,n} \alpha_i |\phi_i\rangle}_{\forall \alpha_i \in \mathbb{R}}$$

The outline of the reinforcement learning constructs that follows Markov decision property are as follows:

Agent: Primarily, an agent acts as an enactor i.e., executing actions throughout an information environment. Here, the reinforcement learning agent as forager (information seeker) explores the search environment inducing textual queries as actions, where the action selection follows via the Actor-critic policy network.

Action: In Web search, an action a_t emanates when searchers manifest their information need via a query in order to locate the document as one of the relevant results or sustain the search process (such as exploratory search). A mathematical description of user action can be depicted in Table 1.2. An action of the forager aims to match a candidate query denoted by $|q\rangle$ (procreated after having a set of input queries) from a document D to depict $|q_{rD}\rangle$, which corresponds to a query state vector as an optimal query for the chosen document D with respect to a positive reward r . The procreated output from the Actor network is a candidate query provided a set of input queries by the forager.

State: The state in RL is analogous to feature representation in machine learning, where a state encodes action in the given information environment and feature representation encodes patterns from the data. The state component possesses a set of actions denoted by s_t that describes the positive historical interaction of a forager in the search environment. The policy function employed in our RL based framework, where the Actor network has its state encoded via the coefficient of global-local projection i.e., the product of their corresponding probability amplitudes $\langle \psi_q^T || \psi_q \rangle$. This projection depicts the word meanings of a query (made up of multiple words). The superscript T in the global-local projection denotes the tensor product. Also, such a state represented is described using the product pooling approach.

State Transition: In MDP, the transition between states depends on the foragers' action embarked at each time step. The state transition can be estimated based on the user's feedback. The state transition estimation employs a convolutional neural network (CNN), where its convolution is made up of state vectors to encode the historical interaction of a forager in finding the match of a query.

Policy: The notion of policy function is to decide which action of forager to divest in each state in order to maximise certain reward functions. We adopt a stochastic policy network, known as the Actor-Critic method [132]. This network specifies which action of a forager is an optimal policy based on the feedback (reward for the corresponding taken action) procreated by the Critic part of the network. Hence, the Actor network depicts the probability of an action by a forager, and the Critic network generates the optimal value (or reward) and renovates it. The policy function represents a probability distribution over the forager's actions and so it is stochastic.

Reward Function: The role of reward $r(s, a)$ is to decide the success value of an RL agent's action in a certain state s . The success value is analogous to the relevance judgement score in information retrieval [205]. In our qRL framework, the reward values are generated from the Critic network based on the input as a pair of state and action, and this procreated value is fed back to the Actor network as an optimal reward for a given action. Then, the Actor network decides and scores the agent (or forager) optimal action.

Based on the RL constructs described above, we illustrate how the proposed framework uses these constructs by encapsulating the mathematics of quantum probability to strengthen the capability of an RL agent to guide information seekers (or foragers) along the search trail in an IR task.

First and foremost, the RL agent is based on the Actor-critic network [132] which is described component-wise as follows:

Actor Network

This is one of the components of a stochastic policy gradient method, in which a forager (or information seeker) inputs textual queries (or forager's actions) under a certain state $|s_t\rangle$ and the Actor network outputs an action $|a_t\rangle$. The input queries are in the form of the sequence $|q_1\rangle, |q_2\rangle, \dots, |q_n\rangle$, which is to capture queries in a search session and it follows the local and global representations with respect to a particular textual description. Such forms of representations are required to model the interrelated interaction between words in an input query. The representational capability of a query is powered by the quantum theory and the foundation of it is based on the necessity of word positions [219], and the interpretation of wave function $|\psi\rangle$ as a state vector can be explicated in reinforcement learning constructs.

The state vectors (or query state vectors) are input to the Actor network and a state vector composes different words in which, each and every word is considered as a state vector denoted by $|w_i\rangle$. Each word has a corresponding unique basis vector $|\phi_{b_i}\rangle$ which delineates

a peculiar semantic meaning via their probability amplitude. The uniqueness of a basis vector lies in the fact that it can lead to a varied meaning if elucidated severally across state vectors (or words), i.e. a basis vector $|\phi_{b_1}\rangle$ refers to ‘date’ in word vectors, $|w_1\rangle$ as ‘event date’ and $|w_2\rangle$ as ‘date fruits’ delineates different meaning when combined with another basis vector. Then, we employ our proposed framework in a query matching task in which a query representation (comprises of multiple words) is of a real-valued vector, where the versatility of discrete basis vectors expands via the local and global distribution of queries. These discrete basis vectors characterise the interaction between the meaning of varied words. Therefore, a query state vector $|q_i\rangle$ can be described through the tensor product of words as $|\psi_q^T\rangle = |w_1\rangle \otimes |w_2\rangle \otimes \dots \otimes |w_n\rangle$. Also, this illustrates the dependency among words by means of a tensor product as

$$|\psi_q^T\rangle = \underbrace{\sum_{b_1, b_2, \dots, b_n=1}^k}_{\text{Probability amplitude}} \underbrace{\mathcal{L}_{b_1, \dots, b_n}}_{\text{Probability amplitude}} |\phi_{b_1}\rangle \otimes \dots \otimes |\phi_{b_n}\rangle$$

$$\text{where } \mathcal{L}_{b_1, \dots, b_n} = \prod_{i=1}^n \alpha_{i, b_i}$$

\mathcal{L} represents the respective probability amplitude and is of k^n dimensional tensor. The associated basis vectors $|\phi_{b_i}\rangle$ depict the meaning corresponding query and \mathcal{L} depicts the local distribution of a query [249] which is of rank 1 tensor. However, words that are unseen in a query (or compound meanings) can not be inferred from local representation. Thus, a set of basis vectors (or states) is required for the global representation, where a query state vector can be described as

$$|\psi_q\rangle = \underbrace{\sum_{b_1, \dots, b_n=1}^k}_{\text{Probability amplitude}} \underbrace{\mathcal{G}_{b_1 \dots b_n}}_{\text{Probability amplitude}} |\phi_{b_1}\rangle \otimes \dots \otimes |\phi_{b_n}\rangle$$

$$\text{where } \mathcal{G}_{b_1 \dots b_n} = \sum_{r=1}^R w_r \cdot e_{r,1} \otimes e_{r,2} \otimes \dots \otimes e_{r,n}$$

The description of the above equations is reported in Table 1.2. This global representation of a query elicits a semantic embedding space of n uncertain word meanings. The distinction among the local and global representations lies in their corresponding probability amplitudes (\mathcal{L} and \mathcal{G}), where queries delineating in global distribution will be trained on a larger corpus (containing historical queries) and queries in the local distribution refer to the input query. The probability amplitudes among words from the local (input query) and global distribution

(unseen words) are computed using the inner product of both representations i.e. $\langle \psi_q^T | | \psi_q \rangle$ which disentangle the interaction. A mathematical description of the project can be found in Table 1.2. This parameterised version of the Actor network uses a convolutional neural network to learn the generated high-dimensional tensor \mathcal{G} , where tensor rank decomposition can be employed (instead of other approaches such as generalised singular value decomposition) to split it and the disintegrated unit vector $e_{r,n}$ with each rank 1 tensor of weight coefficient w_r . These vectors $e_{r,n}$ are k -dimensional unit vectors and it enact as a subspace of \mathcal{G} . The input to CNN is a query state vector with a convolution filter made up of the inner product among $|q\rangle$ and the disintegrated vector and so it makes the CNN trainable. Then, the representation of the Actor's state follows the product of all mapped unit vectors (from \mathcal{G} for all sub-words of a query. Incorporating the network with quantum probability constructs, the Actor network produces an action state vector $|a_t\rangle$ as an output to characterise a set of matched words.

Critic Network

This component of the policy function uses the quantum language model to parameterise the CNN and acts as a discriminator for the Actor network due to the inherent reward as an output. The input to the Critic network is a pair of generated states and the candidate action $|a_t\rangle$ from the Actor network. The Critic network outputs the reward value (a scalar value or Q-function's value [204]). The reward values $R_e \in [-1, 0, 1]$ manifest the merit of the candidate action generated by the Actor network. The importance of reward value delineates the probability of assigning a correct label to the candidate's action. In other words, it is the multi-class classification of queries to match among documents that will be used to update the reward. The reward labels are -1, 0, and 1, where -1 represents a mis-matched query that has negative word polarity (in a compound meaning). Let's consider an example of a query 'dogs chase cats' and 'dogs do not chase cats' that give rise to a compound meaning but in an opposite cognition. We assume that a word renders the overall polarity of a query given to which the new word it superimposes with. The hypothesis can be interpreted with one of the realistic examples with respect to our framework's key constructs - $|q\rangle$, which is a query state vector equable to the tensor products of available words. The coefficients of a word which are the probability amplitudes of basis vectors can be transformed to deduce a new query leading to a compound meaning. This scenario reflects the notion of negative word polarity. The other two reward values 0 and 1 are categorised as a partial match and perfect match for queries.

The Critic component acts as a discriminator, where the concatenation of the Actor's state and candidate action uses one-hot encoding provided the query is fed through a complex-

valued lookup table. Each and every word in the complex-valued lookup table is in its own superposition state and encoded as a complex embedding vector [124]. Then, the complex embedding vectors are transformed into the query density matrix through measurement using the square projection. The probability of a measurement can be estimated via Born's postulate for the given query state ρ (density matrix) which is $p = \text{Tr}(P\rho)$, where p , P , and Tr depict the class of the query, projection matrix, and trace of the matrix. The query's density state $\rho = \sum_{i=1}^n \beta_i |w_i\rangle \langle w_i|$ characterises as the word states in conjunction, given that the density matrix $|w_i\rangle \langle w_i|$ contemplates a word w_i in superposition state (i.e., $\sum_{i=1}^n \beta_i = 1$). The procreated query density matrix has diagonal and non-diagonal entries are real-valued and complex non-zero values, and both type of entries instinctively signifies the distribution of contingent (the diagonal elements represents probabilities of estimating the word states) and semantic meanings (the aforementioned density matrix containing a word w_i is in superposition state between the basis elements that chose to delineate ρ). Also, inspired by the notion of complex phase described in [124] is incorporated in our framework to estimate the sentence density matrix which has word cognition - positive, neutral, and negative. The computed reward values from the Critic network are elicited based on the elucidation from the measurement matrix.

To summarise the conjoined Actor-critic network which resembles a generator-discriminator provides similar settings as in traditional reinforcement learning with respect to a number of components for the constructs. Also, the RL agent in the framework is a controller to the information seeker (or forager) and is analogous to a searcher via the lens of Information Foraging. IFT characterises a searcher an optimal foraging path through information scent and depends on the user to choose an information patch directed via weak or strong information scent. In our proposed framework, the Critic network signals the Actor with an optimal reward value for a certain candidate action which is positively rewarded. Thus, our notion of the quantum probability parameterised reinforcement learning framework meets IFT in definite regards such as the information seeking behaviour of a searcher assessed as foraging and implicitly as RL task.

Reward: The aim of a forager is to identify the perfect match of a query for the chosen or clicked document that is perceived as its reward. However, the proposed framework's reward function is structured in a manner that guides foragers' on how to perceive the document and draw the most relevant match (information patch). Also, the discreteness in the reward value (-1,0, and 1) resembles the concept of relevance in information retrieval. Based on traditional reinforcement learning [204], rewards can be normalised to procreate results. However, reward harmonises a definite analogy of information scent, which is to measure

utility and outcomes in two different types of information scent score - a scalar value and the probability distribution of scent patterns [99]. If this analogy is interpreted in RL, the distribution of attained reward value by an agent can delineate the nature of information scent patterns. Thus, an explainable technique for rewards using the Information Foraging based model of information scent may emulate further intuition to negative rewards. Information scent can be rendered as the perceived relevance of rewarded actions assessed as positive and negative scent values. The physical meaning of positive and negative rewards is that the forager accumulates rich information along the foraged path to find the relevant information, which signals a strong information scent, whereas the unhealthy consumption of information deduces a forager towards an irrelevant patch. This leads to weak information scent and so the forager gives up the information environment (or RL environment) resulting in stopping the search.

Policy and Probability Amplitude Updates: The origination of probability amplitudes in the policy network - Actor-critic is due to the preface of constructs from quantum probability. However, these probability amplitudes need to be updated corresponding to a particular action. The essential excerpt is to measure the forager actions for some definite states, which on collapse will lead to the occurrence probability of the norm of state vector for the peculiar candidate action, that subsequent will execute the Actor network. The importance of the sequence of queries is to capture the forager's experience and learn each of their action (including erroneous actions), the probability amplitude becomes much more informative. Having the forager's action $|a_t\rangle$ as a tensor product of all potential words, one can compute a single action of the forager ($|a\rangle$) during interaction with alteration in probability amplitudes for the combined meaning.

For instance, in Figure 3.3, an illustration of the action $|a_s^{(q)}\rangle$ is the superposition of 2^q possible actions, and to compute $|a\rangle$ can be certain while interacting with changes in probability amplitudes for a quantum system.

Here a single user action (in a certain state) learns in Figure 3.3, where $|a\rangle$ is an action and we compute the weighted superposition of all actions from [75].

$$|a_0^{(q)}\rangle = \frac{1}{\sqrt{2^q}} \sum_{a=00\dots0}^{\overbrace{11\dots1}^q} |a\rangle \quad (3.1)$$

where $|a^\perp\rangle$ depicts an arbitrary state orthogonal to $|a\rangle$, and U_a is a unitary transformation for geometric interpretation which flips the action sign $|a\rangle$ and is trivial on any other action orthogonal to $|a\rangle$ in a 2^n -dimensional Hilbert space. Actually, U_a represents the vector about

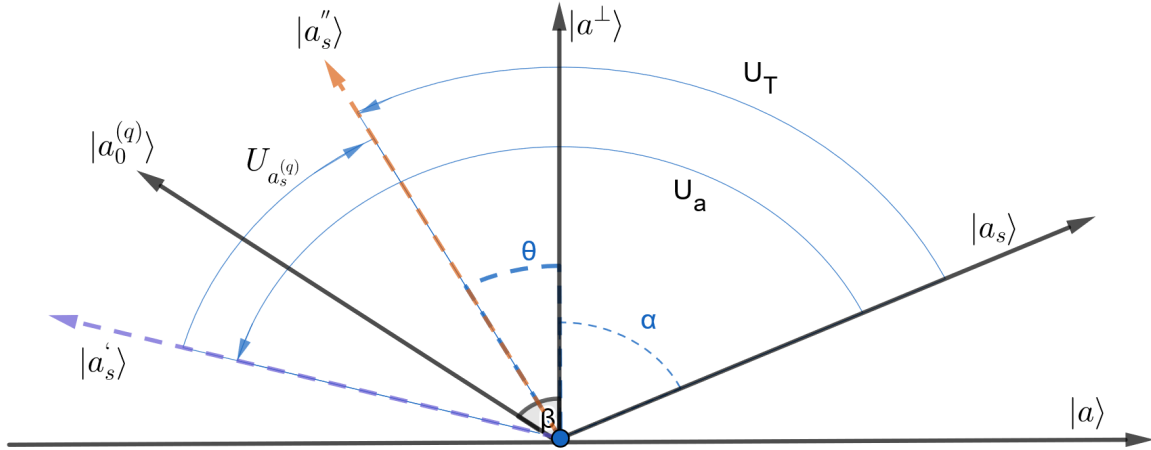


Fig. 3.3 Pictorial representation of a single action and its corresponding unitary transformation for quantum state update in a Hilbert space

the hyper-plane orthogonal to $|a\rangle$, or the action analysis in Hilbert realised to a black-box (i.e., U_a), which can significantly account for whether the given action is an appropriate action. Similarly, $U_{a_s^{(q)}}$ takes on $|a_0^{(q)}\rangle$ and it changes the vector sign orthogonal to $|a_0^{(q)}\rangle$.

3.2.4 Conclusion

We propose a theoretical framework to guide information seekers in a typical information retrieval task of query matching. The presented framework uses quantum theory (in particular, Hilbert space formalism) to encapsulate contextual information and user-level aspects based on the Information Foraging theory. The notion of parameterising the reinforcement learning constructs via quantum probability is to unify the capability of an agent in an information seeking session. The framework models the learning process of the forager's actions in a semantic query matching provided the information environment is patchy. The main importance is to characterise a forager with very little or unclear idea about their information need, vague or evolving information need. Also, a forager has no prior information on how to make their trail (at the start, the information scent is almost zero and emanates as it follows through specific cues) preference during locating information and the amount of information to be consumed in real-time interaction with the search system. Our framework also minimises the primary trade-off of exploration and exploitation by using the RL based policy function - the Actor-critic method. The Actor network is successively informed of

their generated candidate action via the feedback from the critic network in terms of reward. The usage of quantum probability constructs enhances the representation of foragers' search actions and states. However, the implications behind the posed query state vector delineated via a tensor product of words are in practical implementations, especially, dealing with the sparse high-dimensional matrix, and should be fairly computationally expensive. Although, our framework relies on an interpretable theoretical model (qRL) that is inherently able to represent the user's actions. In future, we intend to implement our framework with the *Tensorly*⁶ library. The Actor-critic policy function addresses the issue of continual update of foragers' state-action in parallel via learning and representation mechanisms. Several complex information retrieval problems could be expounded suitably in a new manner within such an inclusive framework.

⁶<http://tensorly.org/stable/modules/generated/tensorly.solve.html>

Chapter 4

Modelling Searcher Preferences in Content-based Image Recommendation based on Information Foraging Theory

4.1 Overview

People usually favour the Web for their everyday searches. The retrieved contents such as documents, images, etc. are primarily the sequel of extensive user searches. In the current era of Web search, in general, users' searching method is based on keywords. However, the increasing functionality such as recommendations via tags, and auto-suggestions are available in almost every commercial search engine. These types of recommendations are frequently used by the searchers depending on their initial information needs. A typical scenario of recommendation can be based on user profiles or the similarity of documents. Recent advances in renovating search effectiveness have a growing interest in user-oriented or personalised search [54]. Personalised search systems apprehend user preferences obtained from explicit feedback (user feedback), which is decisive in web search and image recommendation. In image search, generally, users are reliant on those images to be relevant which they attend during a search. The perception [227] of a searcher characterises an image when they encounter it in the search results (such as search engine result pages (SERPs) which impersonate retrieved images to be hovered or clicked), that is an image to be manifested by a user in their cognitive beliefs. This gives rise to the steady conscious circulation of information by the user and embraces them toward information overload [202, 114]. This phenomenon enables the user search ability to be distant from interesting items (such as images). Also, a searcher deems certain images based on the images attended prior to one

of their searches. The usual information overload [202, 114] in image search/recommender systems has been investigated from the viewpoint of improving the search interface [234]. Image recommendation [141] heavily uses metadata and procreated visual features via deep neural networks to recommend art. SERPs characterise textual and visual information and the user can assess in the context of a seminal state under their foraging behaviour [166, 42, 71]. Information Foraging Theory [166] stems from cognitive psychology to guide users to follow a finite path, navigate, situate and consume information in the search environment. One of the IFT's main constructs - information scent used to frame up SERPs expressly through calibrating alteration in search behaviour. IFT underlay strategies for the user to seek out information (or information patch) that has the strong information scent [240] and conversely explore the available information patches which have weak information scent. The strength of information scent is computed through textual and visual cues from the search environment and reflects the cue's relevance to the search task. This lead to the fact that users generally spend less effort during information seeking and apprehending to maximise their information gain. In text-based recommender systems, users during interaction with suggested items are altered based on their preferences to personalise the recommendation list. The so procreated user preferences are employed to amplify their visual attention for personalised image recommendations. We infer the notion of foraging intervention for explainable (image) recommendation, which is a task of finding the correct item from an enumerated list of recommended images and conferring such interventions can apparently shape the information scent efficacy of user preferences. Thus, the correct foraging strategy assists user's in inferring their preferences and also minimises their cognitive load.

To define user attention within the list of recommended items (images) and the effects of making a choice among the items list, we investigate the impact of adjoined visual bookmarks on images through estimation of the image's information scent.

The main contributions of this chapter are as follows:

- We propose a personalised content-based image recommender system that consolidates users' visual attention to recommended items, and the role of information scent in user-oriented aspects.
- We present implicit feedback signals by means of information scent artifacts in image recommendation.
- We report improvements in the impact of visual cues through information scent.
- We demonstrate the evaluation of the content-based image recommendation system on the collected image data from a visual discovery engine.

4.2 Related Work

Information Foraging Theory has been used to illustrate the impact of image search experience in an adaptive user interaction framework [130]. This framework adheres to quantitative analyses of three enlisted evaluations for content-based image retrieval with varied user types. Also, this framework introduced a user classification model to characterise their interaction with content-based image retrieval and elicits the model on a diverse set of interaction features accumulated based on the screen capture of varied user task types. They validated the user classification model through qualitative data analysis and established that the six different characteristics in the model are compatible with those interaction features. This led to incipient training for studying user interaction or behaviour using IFT.

Personalisation in image recommender system [82, 41] paves the way for different kinds of models such as image-aware and image-unaware recommendations, in particular on social discovery data. These recommender models constructively pull out images from a large collection of candidates that propitiate users' preferences. Image-aware recommender system [82, 41] uses multi-dimension representations and user modelling, where image representation eloquently and distinctively has become one of the effectors for recommendation. [82] introduced a model that leverages pre-trained deep nets, which is selective to embeddings i.e., visual semantic embeddings from pixels matrices (image's pixels). This technique considers an image entirely as a single object, which makes the visual features of embedded objects within the image noisy and difficult to distinguish at a fine-grained level. [41] proposed an attention-aware model for multimedia recommendation in which, the attention network on each side of the model captures image segments with relative usefulness. The de facto standard of this method is that it partitions images into equal-sized regions with the exclusion of semantic objects. We infer that the inclusion of semantic image objects contribute to user preference and on the contrary, exclusion of semantic object results in fallacy for image selection, and so retrogressing the overall effectiveness of image recommendation.

Another type of personalised recommender system is the image-unaware recommendation. This recommendation system relies on user modelling instead of encapsulating images' visual features. The importance of visual features of images can be compared with [175], which proposed a pairwise learning algorithm with implicit feedback for a recommender system, where user-item interaction was explored without image information. However, some of the methods are developed as primitives for patterns in user behaviour/profiles in order to improve the performance of recommender systems [104, 106, 127]. Similar recommender systems are developed with varied approaches such as [185], which proposed a task-oriented (topic) model to inscribe social network effects in a personalised image recommendation.

The perspective of Information Foraging Theory in image recommendation is to characterise the meta-elements (tags, visual cues, or visual bookmarks) in user-item interaction by means of inherent implicit feedback or explicit user preferences. Earlier IFT has been immensely explored on textual data to describe or represent users' information need and their evolving or impeccable actions using information scent [42]. The constructs of IFT are not limited to a single aspect of user behaviour but are also prone to examine and foreshow usability of Web contents through their information scent score [43]. One of the recent works employs IFT to understand the impact of feedback data in a typical movie recommender system, where the alteration in user interface uses information scent at the expense of information access [189]. They established that the prime task of movie selection to watch renovates the feedback data.

Implicit feedback in recommendation systems has been a trend lately as user-procreated contents such as tags, and comments in online forums (such as Reddit, Quora, etc.) were harnessed in varied ways, for instance, to renovate user and item profiles or explanations to the user about recommended items [101, 212]. However, there is a diverse set of information that the recommender system processes based on tags provided by users, this information comprises user opinions, semantics, and sentiments. Additionally, explanations of recommendations to the user are heavily operated by tag-based recommender systems, which depend on user-procreated content and leverage this information to amplify the recommendation quality. Having so, users still encounter stiff competition with recommender systems to effectively avail multimedia content. From the viewpoint of users', [26] investigated that interfaces in textual search lack accessibility to ensue cognitive skills of users. In brief, multimedia recommendation depends mostly on users' explicit and implicit feedback. Therefore, the significance of feature extraction from multimedia contents [83] for the recommendation, and how these features can incorporate with user preference / generated content to endow the relevant recommendations. Implicit feedback underlies observed behaviours exhibited by the user. [153] found user behaviours to be classed under four categories which are retained, examined, referenced, and annotated. Their focus on these classed user behaviours are recommendation system and proposed a classification framework for comparison among users' explicit and implicit feedback with a set of distinct properties. They leverage the classification model for delineating user preferences in items.

This chapter investigates a personalised content-based image recommendation based on Information Foraging Theory, in which we transmute the image items in a representation space to recommend alike items. Also, the users' attention is on examining property of observed behaviours.

4.3 Personalised Content-based Image Recommendation

In an image recommendation system, users are the main actor in making a selection of recommended items. Considering the user viewpoint, we develop a personalised content-based image recommender system (PCBIR) which is framed as *User-Image-Cue* model. The User-Image-Cue model is formulated as an interlinked graph consisting of the user, image, and re-ranked cues set. A pictorial representation of which to demonstrate the system architecture is shown in Fig. 4.1. The precedence of using content-based recommendation

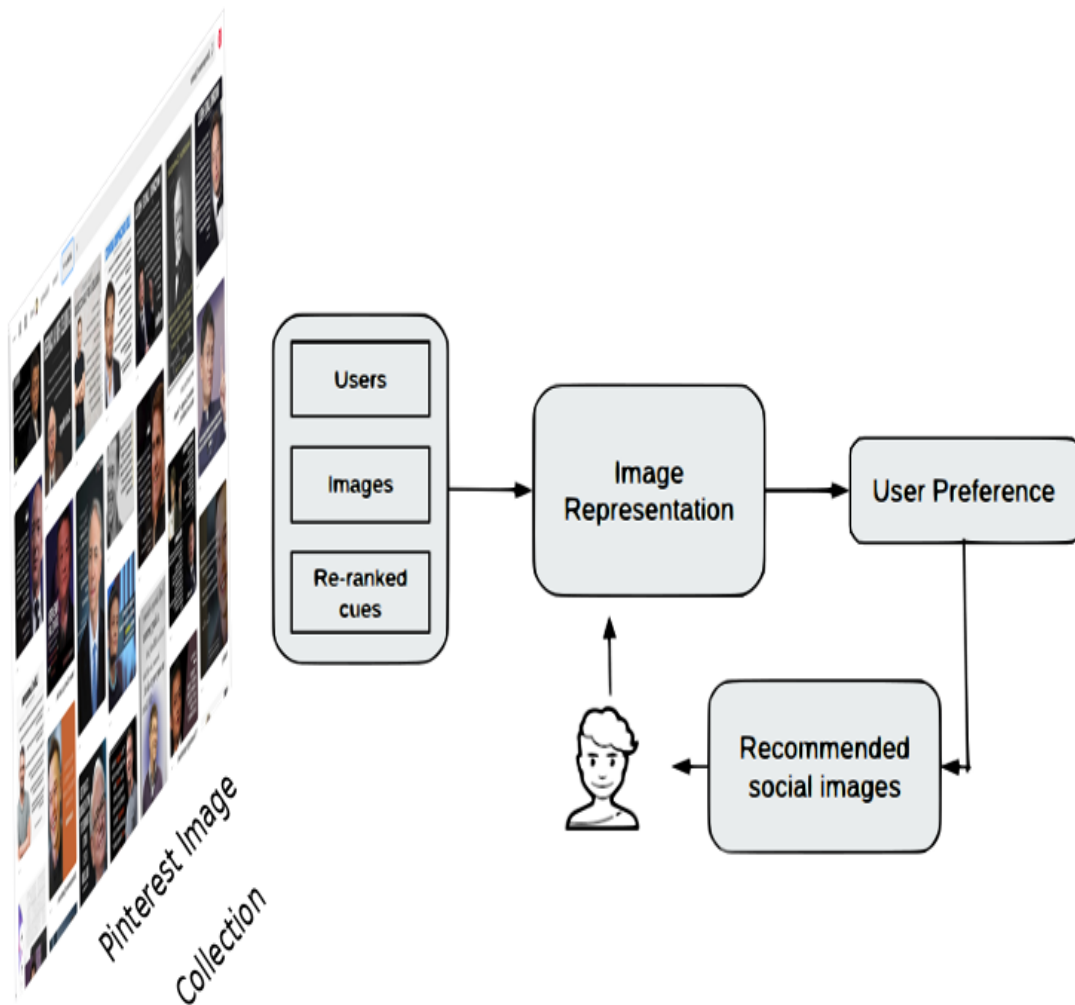


Fig. 4.1 Schematic Architecture of Personalised Recommender System

over collaborative filtering is due to the fact that the latter option is prone to cold start problem [128], where a new item (or user) is introduced without prior history as well as an ample amount of data, whereas the latter leverages the users' correlation to make a

recommendation. The content-based recommendation engine directs the image objects in the representation space, which allows for recommending similar items. Following the comparison among both types of recommendation, we highlight and describe the proposed personalised recommender system reported in Fig. 4.2. Our prototype comprises four screenshots arranged in order from left to right on the top and bottom sequentially.

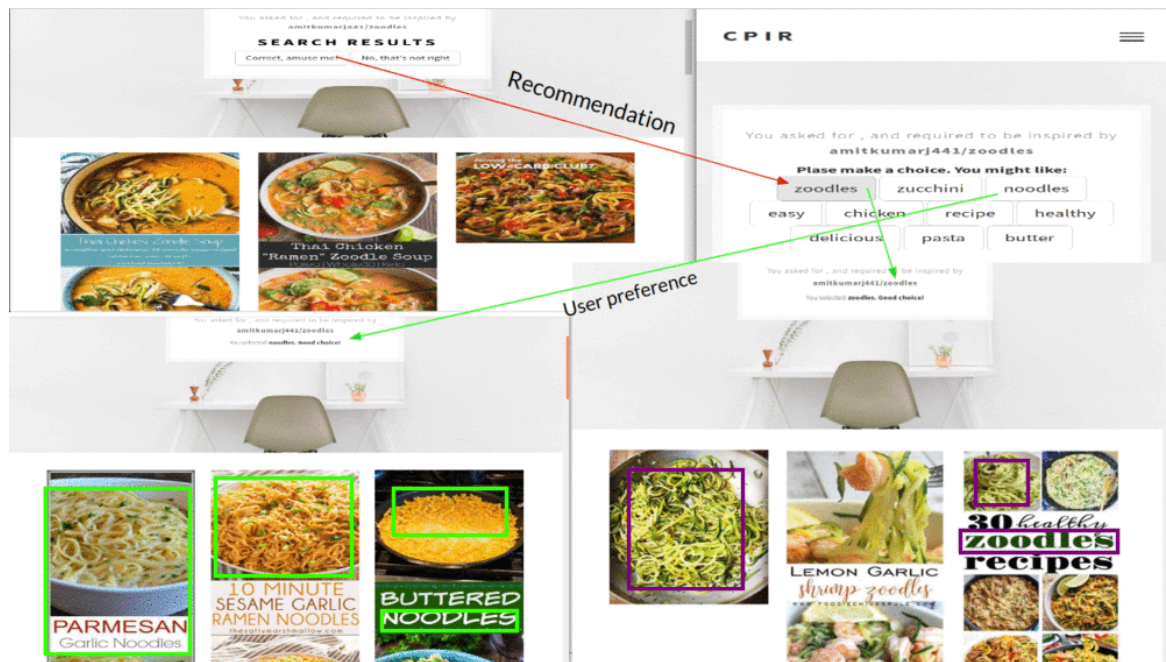


Fig. 4.2 User Interface of Content-based Image Recommender System

The image recommendation system follows the Pinterest components in a way that it can be adoptable to the user and can cater to the user preferences in a simplistic manner. The developed prototype is a web-based application served locally as shown in Fig. 4.2. The first part of the figure on the left consist of a user input field that has a widget of Pinterest¹ board configured in the image recommendation user interface (UI). It asks us to input their board name² in the form of a keyword query. Then, once the user inputs the query, the image recommender system fetches the entire set of images in real time from the prescribed Pinterest board. The second image on the top right is the next involved step, where users follow through with varied recommended preferences. On selecting any of the preferences from the given recommended list, the system retrieves a ranked list of identical items (shown

¹A visual discovery search engine at scale

²It is a collection box that contains a list of images belonging to one or more classes. This is identical to the Pinterest board, which enables the user to establish their varied interests, skills, and plans in the form of visual cues.

via coloured green arrows on the bottom left and right). Each and every recommended images are subscribed with corresponding visual cues.

Image Representation

The representation of an image is with their hard-coded features so as to minimise computation resources during training of image classifier including a simplistic rover of image embedding [62]. It is alike word embedding (word2vec), where images are considered as an {image, label} pair for training using a neural network that transforms the image matrix representation ($224 \times 224 \times 3$) to a much smaller vector representation (image2vec). This technique is employed to estimate the image similarity (nearby vectors delineate alike images) among different images.

Our elucidation to characterise the behavioural aspects of interactive elements (tags, cues, etc.) in an image recommendation system as opposed to spotlighting on designing a unified personalised recommendation system, which is more likely to rely on learning algorithms or users' feature-driven instead of incorporating explainability in such system.

The proposed recommender system depicts various image features such as texture, colour, content, and title/description to characterise. We use a pre-trained ImageNet classifier for each of the features to train on and used each of them on images to elicit the connected information. Also, we then employ ResNet50 – deep residual neural network [81] to train a content classifier on ImageNet³ which identifies nearly 1,000 varied objects. To predict the image colour, we use unsupervised K-means clustering to match predominate colours to the generic colour labels using an HTML colour scheme.

4.4 Experiment I - Content-based Image Recommendation using Information Foraging-driven Interventions

We hypothesise based on the notion of visual Information Foraging theory [167] to locate information more quickly if there exists a strong information scent [42], which relies on cognitive perspectives. We apply visual Information Foraging theory in a personalised content-based image recommendation system to explore, whether the users' attention to a specifically recommended item (i.e., image) exhibits foraging behaviour or can be enhanced with the help of IFT.

³<https://pjreddie.com/darknet/imagenet/#resnet50>

4.4.1 Users' Attention based on Visual Information Foraging

We describe the foraging behaviour in an image search context that spreads implicit behavioural signal (or implicit feedback) for a content-based image recommendation. To investigate the personalised recommendation system for searching images via Information Foraging theory, we delineate this image recommendation system in which the recommended items (or result as SERP – search engine result page) can be viewed as information patch along with all possible image perceived as within a topological frame. The user goal is to find an interesting item as an optimal decision during the foraging loop [164]. We consider images as an exemplary collection of a finite number of patches⁴ which can be attended through cues whilst seeking an image content once user cognitive beliefs [227] activates. A user can have its' cognitive beliefs through generating implicit cues to discern ideas and plans for seeking, gathering, and information consumption. It is analogous to animals' believe in scents to forage [166] which resembles users' following different kinds of cues in assessing image contents and navigating across patch spaces depending on images' scent. Combining images and tags collectively form cues that conform to an information scent. Cues constitute the information diet and cost of information access which are necessary to be acquired for more information when finding interesting images. This leads to three main factors that can be deliberated via the personalised image recommendation system – the strength of the information scent, the effort allied in making conscious consumption of the image information, and information access cost for further searching information about an image.

The elaborated concept of IFT in an image recommendation task can be described based on their constructs. An image I composes a list of n information patch i.e., $I = \{I_{p_{i,1}}, I_{p_{i,2}}, \dots, I_{p_{i,n}}\}$, provided $I_{p_{i,j}} \forall j \in (1, \dots, n)$. The aim is to locate m information patches of which the user (U) attention is known and share the strong information scent with $I_{p_{i,j}}$. The information scent scores of these information patches are computed based on the frequency of user preferences for peculiar content.

4.4.2 Data

We crawl Pinterest, a visual search engine platform to collect images in real-time. We collect a small set of images containing 1,116 samples categorised into two food classes such as *Spaghetti Bolognese* and *Zoodles*. The collected images include their corresponding Pinterest Pins. Then, we perform data cleaning so as to release the metadata containing image descriptions and titles. The frequency of keywords is computed using Naive Bayes.

⁴Images patches are image regions of a particular image when treated separately

Recommendation	Food Categories	Spaghetti Bolognese		Zoodles	
		User Preferences	IS	User Preferences	IS
R_1		Bolognese	10	Zoodles	9
R_2		Spaghetti	7	Zucchini	8
R_3		Recipe	6	Easy	6
R_4		Sauce	6	Pasta	5
R_5		Easy	3	Chicken	5

Table 4.1 Recommendation Result in terms of information scent scores

We share a tiny sample of the actually collected dataset here⁵ to ensure and abide the Pinterest data release policy⁶.

4.4.3 Results

The evaluation result of user preferences leads to the successful selection of recommended items in the personalised image recommender system. Information scent and recommendation are denoted by ‘IS’ and ‘R’. The list of ordered recommendations (R_1, R_2, R_3, R_4 and R_5) in Table 4.1 based on their corresponding information scent score of user preferences. The rankings of these recommendations are based on information scent, strong information scent depicts a better recommendation. R_i characterises the users’ inferred preferences based on the i -th most liked images (for instance, R_1 has ‘Bolognese’ and ‘Zoodles’) in their corresponding collection of food categories. This signifies that preferences with strong information scent (high score value of information scent) are likely to be recommended first and elicited by the searcher. This foraging-based mechanism to drive users’ attention makes them more likely to adopt visual pins, (visual bookmarks or visual cues) with minimum effort through hovering over recommended images instead of memorising the items themselves (the latter reported in [188]). We adopt a Likert scale of 1 to 10 for the evaluation of the proposed image recommender system, where ‘1’ is the least frequent and ‘10’ is the most frequent. The result is reported in Table 4.1 which lists the recommendation based on the information scent scores.

This approach provides the searcher to defer not consuming any sort of further information diet (in case of memorising either buttons/tags or items).

The notion of our approach can be rendered in the perspective of Information Foraging Theory, where an image having either ‘Zoodles’ or ‘Bolognese’ (in R_1) has a strong informa-

⁵<https://gist.github.com/amitkumarj441/a4e85a89581623dbeca3b901e0c17b0e>

⁶<https://policy.pinterest.com/en/privacy-policy>

tion scent. This means that such an image offered as information patches to the user elicits a relatively large degree of attention whilst likely to consume much information (information diet) and having minimum information access costs, for instance in preconditions of the time spent on search.

4.5 Experiment II - Implicit Feedback in Content-based Image Recommendation using Information Foraging Theory

In this experiment, the notion of IFT is to detect users' preferences via implicit feedback, which follows information scent. We hypothesise that users' whilst seeking, their perception magnifies with visual cues in images as implicit feedback that confers information scent. The artifacts of information scent characterise how visual cues in the images help users in locating the recommended items.

4.5.1 Implicit Feedback based on Information Scent

User conviction for information seeking, collecting, and consumption follows their utility in a search environment [166]. Based on the optimal foraging viewpoint in which animals follow scents to forage is analogous to searchers following different types of cues, eliciting information contents, and navigating across the search environment which is influenced by information scent. Cues serve as 'proximal cues' to determine what emits information scent. This is generally represented through web elements which can be recognised as a visible representation of users' mental beliefs. In other foraging tasks, empowering these mental beliefs, endows the partial awareness of the user access path of information content or utility (cost and benefits) [166].

Images are considered exemplary web elements (information patches) that can be attended via textual/visual cues. Users in their search trail whilst attending certain web contents enable their cognitive beliefs. A user may enable such key beliefs through cues (textual/visual), after perceiving ideas and plans for seeking.

Information Scent Artifacts The information scent can be enhanced if incorporated with visual/textual cues. Cues reflect the internal mental representations and endow an external information scent to realign representations in continual information processing tasks. Cues are often in a nutshell generated meaningfully and prospectively to minimise the information

access cost of continual operations. Eventually, the quality of a cue depends on how efficiently and effectively it helps later information foraging, recommendation, and information retrieval.

We investigate the quality of cues for two different kinds of usage - recommender system and information retrieval [65]. For recommendation systems, cues confer user preferences for an image and procreate the information scent to spot related images [131] interesting to the user. These cues confer partial information to decide on navigation paths in real-time to pursue a search. In information retrieval [43], cues provide information scent to spot the document (e.g. keywords to search) and improves activation [6] that have decayed.

Similar cues tend to have varied information scent elicited by the users. Users can update their cues with realigned knowledge obtained throughout the information seeking process. The overview of our pilot user study to evaluate the proposed PCBIR involves three participants to assess a typical recommendation task. The proposed PCBIR consolidates textual and visual cues for a user to follow the recommended list of items and perform a selection from it (or the recommended list of user preferences) of the interesting image. Users' whilst making a selection of images among the recommended list of items follow a cues trail which is assessed (or ranked) based on their value of information scent. Those images assessed through cues are ranked top in the list of interesting items provided they have a strong information scent, and the rest of the available images are ranked in the decreasing order of their information scent score. The estimation of the information scent score uses the existing technique mentioned in [131].

4.5.2 Data

The datasets employed for evaluating the artifacts of information scent in the proposed recommender system are crawled Pinterest images and the WikiArt [181] collection. The crawled dataset from the Pinterest platform consists of images with their metadata (Pins, descriptions, etc.). We collected 1,116 images that classed into two food categories - *Spaghetti Bolognese* and *Zoodles*. These datasets are also used as a part of our pilot user study that embraces finding recommendations in categories such as 'Spaghetti Bolognese', 'noodle', 'painting', 'sketches', 'landscape' and to name a few. Users' have been ensued with two options, (a) to follow images via cue and (b) to pursue images via user preferences. Each and every participant was given ten iterations to locate q recommendations for each of the given image categories. The placement of recommended items is classed into 'interesting' and 'uninteresting', and the results are averaged from ten iterations based on the score of information scent.

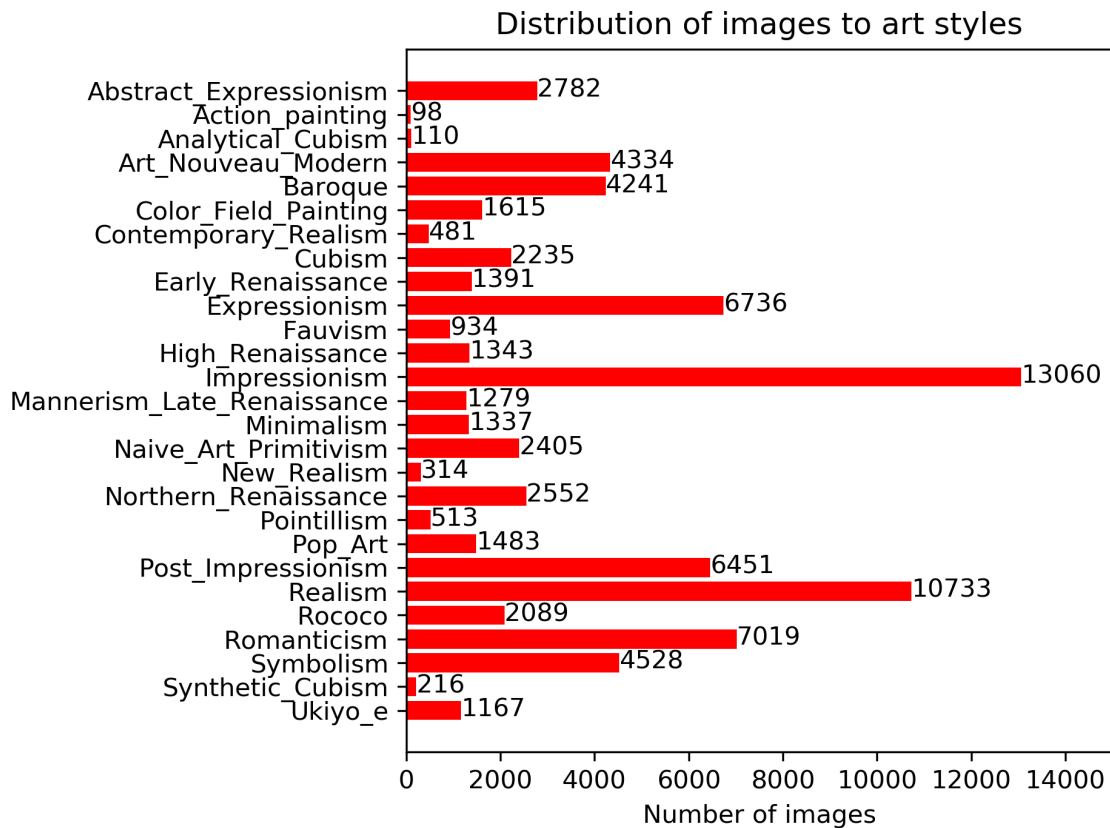


Fig. 4.3 WikiArt Image Dataset

4.5.3 Training

We describe the training settings and steps of the experiment where both Pinterest collection and WikiArt datasets are used. The Pinterest collection is divided into a split of 67% training set and 33% as test set. And, the image labelling is performed on the collected data, where users locate a recommended item and find it interesting to be labelled as ‘1’ and uninteresting as ‘0’, and so this becomes a binary prediction task. The allied information label includes the title and description of the image, which may signify a very complex concept. The cleaning of data follows Naive Bayes analysis to count the frequency of keywords.

We also employ real-world data that has a large number of art images in order to test the generalisation of our proposed information scent strategy. This WikiArt dataset [181] contains artwork images that characterise the images and their diverse features whilst recommendation. The evaluation of our proposed image recommender system follows the WikiArt dataset

consisting of over 80,000 images of artwork labelled across 27 diverse art styles⁷. A list of categorised image classes is shown in Figure 4.3.

We employ the features from the collected image data such as the image content, image size, colour, and aspect ratio to delineate user beliefs during image recommendation. The WikiArt dataset is a large-scale images dataset of art, and we chose only 1,000 image samples as the training set and test set containing nearly 500 images. Each category of images in WikiArt dataset has multiple classes and due to the high diversity among it, we sample images from ten sub-classes of image categories which are ‘abstract_painting’, ‘cityscape’, ‘genre_painting’, ‘illustration’, ‘landscape’, ‘nude_painting’, ‘portrait’, ‘religious_painting’, ‘sketch_and_study’, and ‘still_life’. These ten sub-classes are chosen from 27 varied art image categories. The ten sub-classes are drawn to confer conciseness to the user. For training the image classifier, we employ a pre-trained residual network i.e., ResNet34 [81] to train over these 1,500 images that contain the aforementioned image features. The training steps have parameters involved such as learning rate and epochs, which are set to 1e-3 and 30. The classification among ‘interested’ and ‘uninterested’ classes were based on the accuracy score, where the target labels are the ten sub-classes of images. However, we found only four sub-classes of images (‘abstract_painting’, ‘illustration’, ‘nude_painting’, and ‘still_life’) as ‘interested’ and leads to high classification accuracy (54%, 66%, 65% and 57%). Aforementioned, users’ interesting images are in the provided sub-classes with their corresponding accuracy. Adversely, the users’ uninterested images pertain to the rest of the six sub-classes with their corresponding classification accuracy in order as ‘landscape’: 14%, ‘cityscape’: 28%, ‘religious_painting’: 32%, ‘sketch_and_study’: 35%, ‘genre_painting’: 36%, and ‘portrait’: 37%. The information scent effects are supported by the prediction labels and contemplated during computation of the classification accuracy in a manner, that artwork images retract a considerable amount of effort in selecting the recommended images.

4.5.4 Results

The evaluation result of implicit behavioural signals inferred via information scent scores are compatible with the task of binary prediction, where weak and strong information scents are depicted as ‘0’ and ‘1’ respectively. For evaluation, we fine-tune the scalable vector machine (SVM) and random forest models using grid search (GS) on the Pinterest collection, and XGBoost on WikiArt dataset. The evaluation result is reported in Table 4.2. In Table 4.2, the performance metrics precision and recall are denoted as, ‘P’ and ‘R’. The rank of predictions

⁷[https://github.com/cs-chan/ArtGAN/tree/master/WikiArt Dataset](https://github.com/cs-chan/ArtGAN/tree/master/WikiArt%20Dataset)

Class \ Model	Pinterest Collection						WikiArt Dataset		
	GS-SVM			GS-Random Forest			XGBoost		
	Scores								
	P	R	F1	P	R	F1	P	R	F1
uninterested (0)	0.77	0.85	0.81	0.80	0.89	0.84	0.81	0.62	0.70
interested (1)	0.81	0.70	0.75	0.85	0.74	0.79	0.47	0.53	0.50

Table 4.2 Classification Report

to the rank of correct class labels are measured using the area under the receiver operating characteristic (AUROC) which is shown in Fig. 4.4.

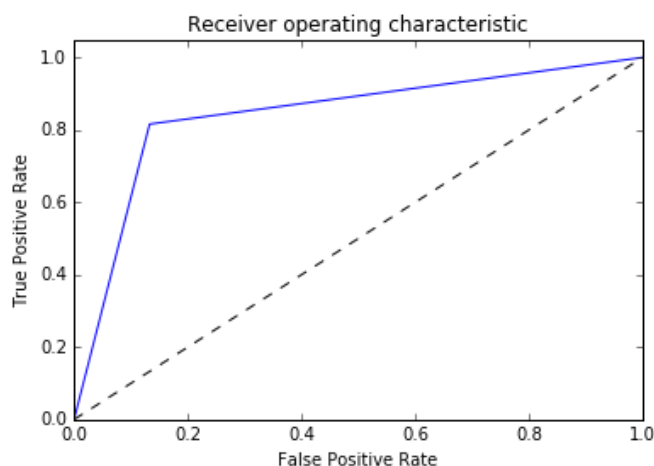


Fig. 4.4 Prediction Performance

Theoretical intuition from the viewpoint of Information Foraging Theory, let's consider an image identified by a user either as 'Zoodles' or 'Bolognese' possesses a strong information scent and so the recommended items entail visual cues. This leads to an image endowed as an information patch (such as rectangles around images in Fig. 4.2, users' whilst following these information patches likely to enrich most information with minimum information access costs, such as in preconditions of the time spent on search. We also place a small set of recommended list of items on Pinterest for Zoodles⁸ and Spaghetti Bolognese⁹.

The employed datasets for evaluation of our image recommender systems are the Pinterest image collection and WikiArt. The list of recommended images accessible to users via textual/visual cues and ranking of the list is based on their corresponding information scent

⁸<https://pin.it/OJ8BFMK>

⁹<https://pin.it/59ow3wx>

score afterward the pilot user study. The enhancement to this work follows by incorporating users' implicit feedback in the proposed recommender system. The validation in terms of performance of the proposed image recommendation system based on Information Foraging Theory is carried out as a classification task. The images employed in the recommender system to see if the inclusion of additional aspects, such as implicit feedback improves the users' interesting and uninteresting items and so is framed as a classification task. The classification result is reported in Table 4.2. Certain observations based on the classification result can be envisaged, first, the F1 score of the 'uninterested' items is significantly higher than the 'interested' items list. This is due to the fact that the diminishing return of information scent score (aligns with the notion mentioned in [15]). This also signifies that the chosen recommended item by a user at first glance interpolates less effort (selection cost), and those selecting user preferences in multiple iterations to choose a recommended item leads to weaker information scent (so it ranks lower in the list of interesting images). Additionally, the poor performance on the WikiArt dataset in terms of F1 score is due to varied multiple colours and non-specific objects in information patches (or images).

4.6 Conclusion

The emergence of behavioural information retrieval frameworks, such as Information Foraging Theory, inspires investigation of how image recommender systems can be enhanced without an algorithmic approach. The IFT-based approach proposed in this chapter has conducted an empirical evaluation of user preferences by means of information scent in a content-based image recommendation. This helps in understanding the impact of user attention in an image recommendation task.

The finding from this preliminary study is that the information scent of an image has user-oriented aspects and the users' information scent of an identical image can differ such as 'Spaghetti' and 'Bolognese'. It has been observed that the entire information scent of an image (such as [131]) magnifies with visual cues. Also, the information scent becomes stronger by reinforcing visual attention but in certain cases, the information scent of images can exceed the cues' scent. So, studying the influence of information scent and cue strength either in form of feedback over each other or as part of the interaction between them, can also be convenient. We consider this as part of our future research directions.

The task of incorporating implicit behavioural signals (features such as presentation context of image content) in an image recommendation set forth our second task. The finding of this task is that the implicit features (Pinterest Pins, visual bookmarks as cues) can be consolidated with images and help shape the recommended items for users'. Also, for the

WikiArt datasets, a subset of it emanates implicit features such as colours and image sizes. This artwork image collection with different image sizes in a large collection is shown to exhibit strong information scent for images with the larger size. It is due to the fact that these images contain diverse scriptures and act as textures that conjunctively renovates the user's perception. Such recommended list of images is likely to be attained by the user. These features assist users in moving the burden of consuming irrelevant information (such as memorising items or buttons/tags). This also resolves the problem of information overload and shapes users' cognitive beliefs whilst locating interesting items.

Chapter 5

User Information Needs in Image Query Auto-Completion based on Information Foraging Theory, using Language Model

5.1 Overview

In general, the user information needs are unspecified and happen to provide a peculiar space for them to refine their input query in the search engine until they find a meaningful or relevant outcome. In query auto-completion (QAC), the information needs of a user is an action (seldom unclear at the start) that evolves and competes with prefixes to form a full query. Users' typing a prefix of a certain character in order to attain a complete query, facilitates the creation of user query [35]. The process of auto-completing a query prefix can also be referred to as (dynamic) suggestion of query [112] and real-time query expansion [230]. One of the prevalent features of information retrieval - is query auto-completion which drives the user to be dependent on search engines to find any relevant information. However, the notion of such a feature prevail users' to express their information needs (via queries) ambiguously, which are then immensely vague to be completely rendered by search engines. This introduces a bottleneck component in the usability of search engines for query auto-completion task [37]. Also, users often apply several rounds of search to reformulate their queries further to adhere to their information needs given they find some relevant results. Past work [42, 137] demonstrated the use of information scent to model users' information needs during a web search, and it has been used to understand the factors affecting search and what takes a user to stop the search. Despite the good observation, the exploitation of information scent (from Information Foraging Theory [166]) is under-explored in case

of ambiguous queries and has not been extended to take into account an image in query expansion (or suggestion) tasks. For the users' convenience, current search engines generally endue query suggestions for them in order to describe their queries more explicitly. They have been explored extensively in query auto-completion tasks, especially the traditional approach known as Most Popular Completion (MPC) [21] which at the extreme is incapable of anticipating a query it has never seen before. Solutions are further improved by recent semantically-driven models [144, 145] and neural model [159] approaches which are the current state-of-the-art in QAC. However, most of the language embedding models [95] have obtained strong results on multiple benchmarks for understanding the polarity of word compositions. Unsupervised pre-trained natural language embeddings [50, 138] successfully model long-term dependencies with the purpose of predicting masked terms and assessing if sentences ensue one another, which showed strong results on several natural language processing and information retrieval tasks. Empirically, recent advances in sequence models have been adapted to span a prefix to full text and index [93] but despite the attainment, it has not been generalised to take an image into account. Also, deep neural networks are mature enough and capable of segmenting regions within an image [80, 87].

To address the above mentioned gaps, we move one step forward to present a method that extends and modifies the state-of-the-art approaches in query completion and text embedding. We apply our ideas to an image search scenario where we assume patches are regions of images that are relevant to the user's information need. Our work is concerned with providing users of image search engines with a useful query suggestion (via a visually-oriented patch form) during an interaction, to further amplify their exploratory search experience. Hence, finding useful patches for query expansion in an image based on textual queries (or descriptions) is the primary focus of our work. Past work [89, 195] used both the query and image for typical retrieval and segmentation tasks. In our task formulation, we rely only upon a given arbitrary text prefix rather than having the entire text query which is used to perform a search based on the image and supported by a modified deep language model [93] to find the most relevant patch in the image. We break down the task into three sub-tasks:

- a) Completing a query from the user query prefix and an image.
- b) Finding patch probabilities based on the complete user query.
- c) Aligning and segmenting all patches in the image.

Our contributions can be summarised as follows:

1. We present a method for image query auto-completion where a user query prefix is adapted upon an image.

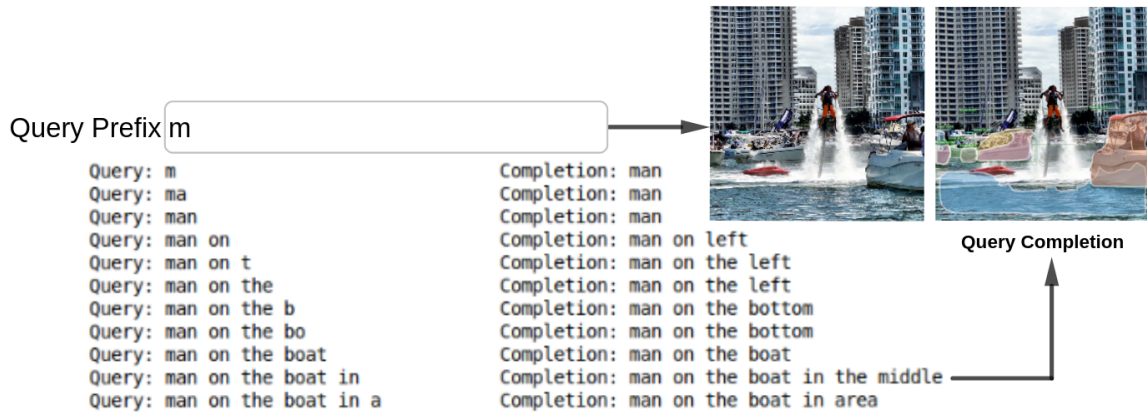


Fig. 5.1 An instance of Image Query auto-completion using our extended LSTM language model

2. We elaborate on the analogy of query auto-completion based on Information Foraging Theory and propose an explainable strategy for the observed challenges of query formulation and the varying users' information needs.
3. We propose iBERT inspired by [50] to compute probabilities of patches and rank them efficiently in the image.

5.2 Related Work

A comprehensive insight on query auto-completion and suggestions in image search, followed by IFT and language embeddings is explicated in this section.

5.2.1 Query Auto-Completion

Query auto-completion is an important aspect of information retrieval systems which allows it to predict what could be the next character (or query item) right after the first key was pressed by a user. The predictions in IR systems are generally driven by the query logs (or query history) which are the factual queries that users have previously entered as they were trying to satisfy their information need [230, 102]. [21] introduced a method called *NearestCompletion* that addresses the situation of 'context' which depicts the users' preceding queries in suggestion-based IR systems. The authors' proposed MPC mechanism relies on the entire popularity of the queries conforming to the provided prefix. Recent work reported in [103] studies user reformulation behaviour by leveraging textual features, whereas [196] introduced personalised query auto-completion and found that utilising a user's long-term

search logs and locations, as well as both context-based textual features and demographic features, is more effective. More recent advances in QAC using neural language models are proposed in [159] using recurrent neural networks that effectuate the performance on immediately unseen queries. A generalised and adaptable language model for personalised QAC is introduced in [93]. We extend this adaptable language model to query completion in an image search scenario in the following section.

5.2.2 Query Suggestion

Query suggestion and query completion differ in their end goal in which the former search aspect outputs a list of ranked queries against an input query, whereas the latter search aspect outputs queries with the first few characters (or text) similar to the user's input. Recent work [239] introduced a learning-based personalised suggestion framework for query suggestion that uses both visual and textual queries. Their work uses users' click-through data. A new paradigm of attention-based mechanisms for referring expressions in image segmentation [195] is proposed which contains a keyword-aware network and query attention model that demonstrates the relationships with various image regions for a given query. Inspired by the idea of attention models, we modify this mechanism for patch alignments within images via information scent in the following section.

5.2.3 Information Foraging Theory

Information Foraging Theory (IFT) [166] is a theoretical framework for understanding information access behaviour, derived from the ecological science concept of optimal foraging theory which applies to how humans access information. IFT stands on three different models, namely the information scent model, information patch model, and information diet model, which can illustrate users' search preferences and behaviours [129]: (1) The information within a certain environment scattered in form of *patches* (images, text snippets, documents) consisting of *information features* (colours, words) refers to the *information patch model*; (2) A user can go from one patch to another via a *cue* (e.g., typing a query by following perceptual or heuristic cues [202]), which meets the user's information need. The goal of such cues is to characterise the contents that will be envisaged by trailing the links, which refers to the *information scent model*; (3) Different types of information sources will vary in their information access costs. Users will assess the information sources based on information gain per unit cost or varied profitability, and then the users will narrow or expand the diversities of information sources based on their profitability. This user behaviour refers to the *information diet model*.

One of the main IFT concepts is *information patches*. For instance, sections and their associated features in search engine results can be considered patches. From a foraging perspective in image search, the searcher is the predator (or forager [237]), the information patch is any segment or a region within an image (or image itself) in a given information environment. The piece of information a user is looking for is the prey, and the consumed (or gained) information is the information diet. Something on the user interface that informs users about a specific place they should look next is referred to as a *cue* of the information scent.

5.2.4 Language Embeddings

Nowadays, many information retrieval or natural language processing tasks rely on language embeddings, such as Word2vec [143], Glove¹, and fastText². They use vector word embeddings for word representation to transform a distinct space of human language into a continuous space, which will be further processed usually through a neural network. In query auto-completion, embeddings have been employed for distributed representation of queries based on a convolutional latent semantic model [144]. Word embeddings have been used to compute query similarity for query auto-completion [193], incorporating the features with the Most Popular Completion model. Very recent work [50] introduced a pre-trained deep language model known as BERT which has shown promising results on several IR and natural language processing tasks. However, it is still not well-explored how to leverage such pre-trained language models for QAC, which poses certain challenges both regarding the task and training. Based on this work, we describe our proposed BERT-based model for computing patch probabilities in the following section.

5.3 Problem Formulation

The task of query auto-completion for image search is formulated as a probability maximisation problem described below.

The image patches in a set $p_k \in P$ provided P depicts the entire set of distinguishable patch classes. The user partial information needs in the form of a prefix explicated via a query q_p as an input, to retrieve the relevant image. A character-based language model is utilised and extended to auto-complete the subsequent query q . This query auto-completion task is devised to maximising the probability of an input query adapted to an image as reported in

¹<https://nlp.stanford.edu/projects/glove/>

²<https://fasttext.cc>

Equation 5.1

$$q_{a^*} = \underset{q}{\operatorname{argmax}} P(q|q_p, I) = \underset{\{t_1 t_2 \dots t_n\}}{\operatorname{argmax}} P(\underbrace{t_1 t_2 \dots t_n}_{\text{terms}} | q_p, I) \quad (5.1)$$

where q_{a^*} depict the adapted query on an image, $t_i \in S$ represent the term in position i within a sequence S .

5.4 Model

Our task of query auto-completion for a text prefix employs image features procreated through pre-trained embeddings of equivalent modality. This is due to the fact the auto-completed query q_{a^*} later fed as an input to the image transformer(modifies BERT [50]) to estimate the image patches³ probabilities, which is considered as a multi-label classification problem.

Assuming the auto-completed query q_{a^*} evinces a set of patches $P_{q_{a^*}}$ in which the estimate of $P(p_k \in P_{q_{a^*}})$ is \hat{q}_{p_k} , and $y_k = \mathbb{1}[p_k \in P_{q_{a^*}}]$. The patch selection model minimises the sigmoid cross-entropy loss function as below:

$$\mathcal{L}_{f_{\text{selection}}} = - \sum_k y_k \log(\hat{q}_{p_k}) + (1 - y_k) \log(1 - \hat{q}_{p_k}) \quad (5.2)$$

The schematic architecture of our proposed QAC pipeline and patch selection model is reported in Figure 5.2. A typical scenario of auto-completion is considered, where the user types their text query prefix for the provided image to auto-complete. The network module contains an image feature extractor that employs a pre-trained Convolutional Neural Network (CNN). These image feature vectors are fed into the extended Long Short-Term Memory (LSTM) language model alongside text query prefix, which is via a context-dependent weight matrix that comprises of adaptation matrix framed from a context-driven embedding model. The corresponding constructs from an image and a text as visual features and textual queries are employed to complete a query prefix. The auto-completed query then as an input to fine-tuned BERT language embedder (iBERT) which computes the patch probabilities, which are harnessed for the selection of patch. A comprehensive insight is in the following section.

5.4.1 Image Query Auto-Completion

User query prefixes with the image features generated from a pre-trained CNN are input to an extended LSTM model (by incorporating a context-dependent weight matrix) which predicts

³Each class of image patches can independently exist

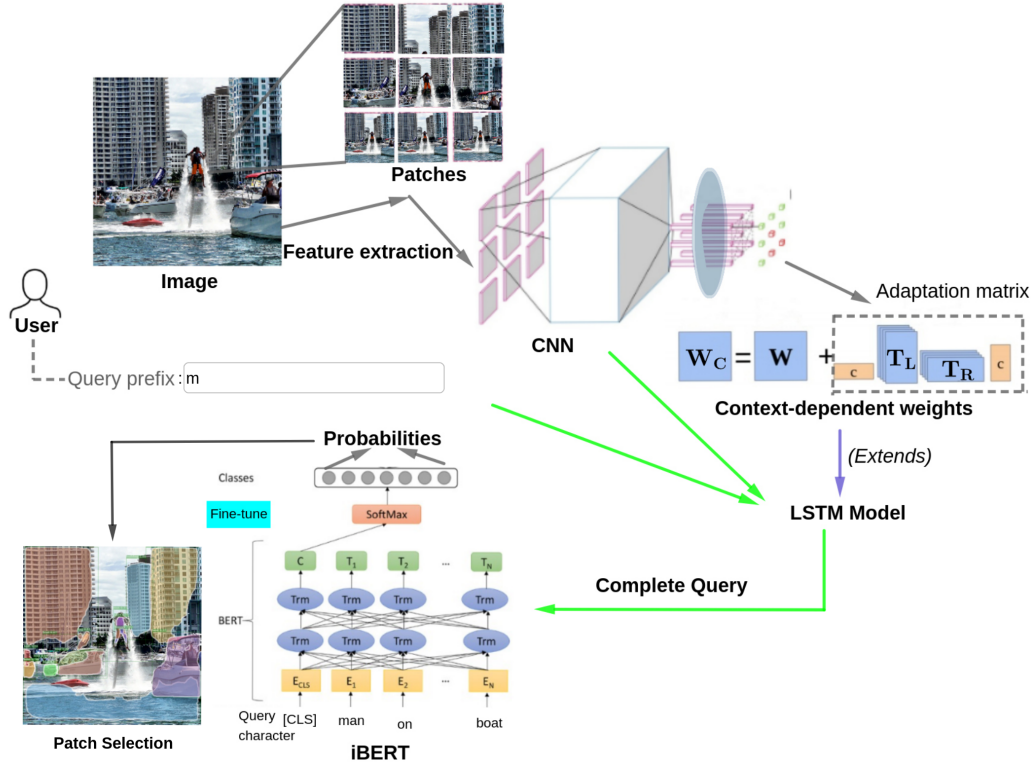


Fig. 5.2 The end-to-end architecture of Image Query Auto-Completion.

a complete query. The resulting query is fed into a fine-tuned BERT pre-trained embedding model which outputs patch probabilities for patch selection. The overall pipeline of the architecture is shown in Figure 5.2. The challenge of query auto-completion is to predict and generate queries from prefixes that have never been seen in the training set. An initial attempt using neural language models has been introduced in [203]. The benefit of using character-level neural language models is providing more fine-grained predictions but they suffer from the semantic understanding that word-level models provide. For a prefix that has not been seen before (such as an incomplete word), their model enriches the shared information among comparable prefixes to create prediction nonetheless. In our scenario, we are given a prefix to complete a query conditioned on an image. To solve this new QAC problem, we exploit and extend the Long Short-Term Memory (LSTM) language model [93] with combined input and forget gates to auto-complete queries. The language model is made up of a single-layer character-level LSTM with layer normalisation [18]. Our extension and modification to this language model are that we replace user embeddings with a low-dimensional representation of images. We adapt this LSTM language model alongside a context-dependent weight matrix \mathbf{W} replaced by $\mathbf{W}_C = \mathbf{W} + \mathbf{M}_A$. We are providing a character embedding $w_c \in \mathbb{R}^e$, a preceding hidden state $h_{c-1} \in \mathbb{R}^h$, where \mathbf{M}_A is the adaptation matrix constructed by the

product (\times_i denotes the i -th-order tensor product) of the context c with two basis tensors, $\mathbf{T}_L \in \mathbb{R}^{u \times (e+h) \times v}$ and $\mathbf{T}_R \in \mathbb{R}^{v \times h \times u}$. Alternatively, the two basis tensors i.e., \mathbf{T}_L and \mathbf{T}_R are re-shaped to $\mathbb{R}^{u \times (v(e+h))}$ and $\mathbb{R}^{vh \times u}$. So the next predicted hidden state and the adaptation matrix can be equated as follows:

$$\begin{aligned} h_c &= \sigma([w_c, h_{c-1}] \mathbf{W}_C + b) \\ \mathbf{M}_A &= (c \times_1 \mathbf{T}_L)(\mathbf{T}_R \times_3 c) \end{aligned} \tag{5.3}$$

We combine the context-driven weight matrix and the immediate preceding hidden state followed by the generated adaptation matrix which is able to alter each query completion to be personalised to a particular image representation. We perform feature extraction on an input image using a Convolutional Neural Network (CNN) trained on ImageNet (pre-trained CNN), where we retrain only the last two fully connected layers shown in Figure 5.2. The generated image feature vector is then fed into the LSTM language model via the adaptation matrix. We leverage beam search decoding [213] in the generated array of predicted characters to select the optimal completion for the user query prefix.

5.4.2 iBERT for Patch Probability

We describe our approach to computing the probability of image patches which addresses an important aspect of query auto-completion systems. We assume that during the search process, users are typically interested in some part of the image as well as the image itself if it matches the mental picture of their belief [227]. Our work focuses on a new perspective of query auto-completion on images and the proposed model finds image patches that match the user context based on the query prefix using equation (5.1). BERT (Bidirectional Encoder Representations from Transformers) [50] shows promising results in multiple tasks of natural language processing and information retrieval [146] and is presently the state-of-the-art embedding model. We propose to fine-tune the BERT model as a transfer learning task for patch selection, using images composed of several patches (regions of an image), hence the name iBERT⁴. To the best of our knowledge, BERT has not yet been retraced for the QAC task. We use the BERT embedding model, which has a twelve-layer implementation, extending it by adding a dense layer with 10% dropout which then is mapped to the final pooled layer connected to the object class, and which outputs patch probabilities as shown in Figure 5.2.

⁴The lowercase ‘i’ represents image patch

5.5 Information Foraging Perspective on Visual Information Needs

Our goal of using Information Foraging Theory [166] from a cognitive viewpoint is to find explanations for the observed behaviour in query auto-completion and to model the information needs within query sessions. IFT postulates that human information seekers follow an information scent to navigate from one information region to another in an information environment that is instinctively patchy in nature, and from one information patch to another within a region. IFT implies that foragers adapt their behaviour to the structure of the information environment in which they prevail such that the entire system (encompassing the information seeker, the information environment, and the interactions among these two) tries to maximise the ratio of the expected value of the information gained to the total cost of the interaction. Following the IFT analogy, when users start typing a prefix to auto-complete, their perceptual cues (such as mental beliefs [227]) either allow them to type the next character or to access the provided suggestion (under the query field) which acts as a distal cue and visually inspires the user to acquire them instantly to forage or seek. Query auto-completion, from an IFT perspective as query-level user interaction, is initiated by the user typing as little as a single-character query prefix. The user then may follow suggestions in case the completion is generated (which again follows the earlier mentioned strategy). In case the query prefix is unknown to the system (e.g. by being entered for the first time) the information scent associated with a result might be too poor [42] to immediately infer information needs. In this case, we are applying beam search to generate the query based on image features. Suggestions are based on information scent values as described in the following subsection. These query suggestions represent the diversity of information scent patterns which elicits a varied distribution of relevant queries in the search field.

5.5.1 Patch Selection

This section describes the foraging-based strategy for patch selection. The technicalities of ranking patches (after patch selection) in the image (from image search results) are illustrated in Section 5.4.2. We utilise IFT to infer the user's information need leveraging the IUNIS algorithm [42] which was proposed to weigh each page vector along with the two factors i.e., TF-IDF weight and time, that were used to quantify the associated information scent with the page. In our image search scenario, we have images as search results where an image is considered as a set of patches containing features such as colour, shape, texture, etc. In our proposed iBERT model, we use information scent to inspect patches based on image

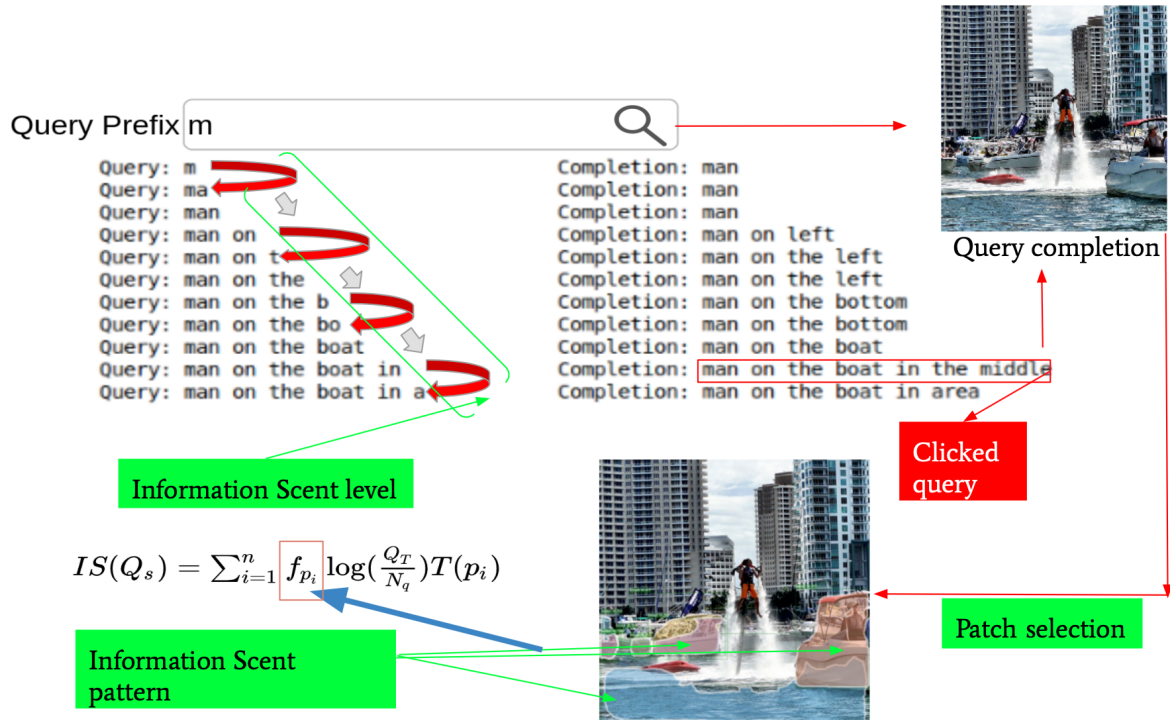


Fig. 5.3 IFT based Explanation of Image Query Auto-Completion

features and select patches which have a higher probability estimated by the iBERT model. *Probabilistic Patch Selection Model* (PPSM) is a first attempt to reflect users' information need coherently by means of information scint. PPSM is used for a task that extends finding patches and makes the quantification of semantic uncertainties an important choice in selection. The important requirement for PPSM is a model (iBERT) that identifies patches in an image that are relevant to the user's information need (query). Inspired by the concept of TF-IDF in IR, we represent the categorical distribution of frequency (f_{p_i}) of each patch in an image (from the search results) in a given query session Q_s and the ratio of the total number of query session (Q_T) during the entire search process to the number of query sessions (N_q) that contain the given patches (p_i) found in Q_s . We also consider the time spent (T) on the resulting images in a given query session to estimate the information scint (IS) within a query session as:

$$IS(Q_s) = \sum_{i=1}^n f_{p_i} \log\left(\frac{Q_T}{N_q}\right) T(p_i). \quad (5.4)$$

The user effort in terms of time is a function of patches which can be diverse and of different image class categories. To generalise this for finding the information scint of a patch which then is assessed to select patches with higher information scint and then compared against the patch probability obtained via iBERT to distinguish the result. If we assume that the generated

auto-completions induce several suggested queries (representing different information needs) simultaneously, every suggestion is in a competition to be discriminated as evident to the user. In the same way, an image contains several related or unrelated patches within it, and users find it difficult to judge which patches are relevant among images, which is due to the high uncertainty of correlated features within an image spread via patches. This motivates us to estimate the information scent of an image patch. There are two ways to compute the information scent of an image; one is to hire individual judges to rate scent on a scale [166] and the second approach is an algorithmic approach [167]. To estimate the information scent of a patch, we consider that PPSM constitutes patches that are probability distributions over images as *observations*. We assume image features as activators to perceptual cues because the user interpretation of image features when matched gives rise to a selection of an item (i.e., patch). The distributions are independent Bernoulli distributions of the features. Each observation is allocated to a patch, but the number of patches is not necessarily fixed i.e., the model is a non-parametric mixture with a product of independent Bernoulli distributions as observation model. Therefore, the log-probability of selecting an image I for patch p_i

$$p(I | p_i) = \prod_{q_p} r_{pf}^{i_f} (1 - r_{pf})^{1-i_f} \quad (5.5)$$

where $r_{pf} = f((\pi_i, s_i), (1 - \pi_i)s_i)$ is the Bernoulli rate for patch p to emit feature f , i_f is the image containing feature f , and r_{pf} is a function of prior parameters representing activators (perceptual cues) for the selected patch. There can be a situation when most patches have only one observation (image) and features are very sparse i.e., the possibility of multiple perceptual cues per patch (i.e., $\pi_s \ll 1$) is low. To interpret Bernoulli's prior parameters such as s_i , we find the probability to observe a feature ($f \in i$ meaning $i = 1$) provided that it has been observed for a patch p ($k = 1$) is:

$$p(i = 1 | k = 1, n = 1) = \frac{s_1 \pi_s + n}{s_1 + n} = \frac{s_1 \pi_s + 1}{s_1 + 1} \approx \frac{1}{s_1 + 1} \quad (5.6)$$

if $\pi_s \ll 1$. The probability of observing a feature in a new image, given that it has been observed before, is a measure of its reliability. We use this probabilistic model to compare the results based on the probabilities of patches obtained from iBERT.

5.6 Experiments

5.6.1 Datasets

We use two well-known and diverse datasets: a visual dataset with large-scale knowledge bases that provide a rich collection of language annotations for visual concepts known as *Visual Genome* [117] with over 100k images where most image categories fall within a long tail, and the *ReferIt* dataset [111] which contains $\sim 42k$ image regions with descriptions. These two datasets fit well for our tasks. The Visual Genome dataset includes images, region descriptions, question-answers, objects, relationships, and attributes. The region descriptions confer a substitution for queries as they refer to several objects in various regions of every image. Few region descriptions are referring phrases and few of them are quite alike to descriptions. For example, referring descriptions are ‘guy sitting on the couch’, ‘white keyboard on the desk’ and non-referring descriptions are ‘couch is brown’ and ‘mouse is in the charger’. The huge number of instances from the Visual Genome dataset makes it quite convenient for our task. The ReferIt dataset is a collection of referring expressions engaged to images that quite intently resemble probable user queries of images. We separately train models for query auto-completion and patch selection using both datasets.

5.6.2 Training

We combine query and image as pairs by utilising the region descriptions from the Visual Genome dataset and referring to expressions from the ReferIt dataset. During training, we have taken 85% of the Visual Genome data as the training set consists of 16,000 images and 740,000 corresponding region descriptions in which there are approximately 40-45 text descriptions per image. The training data from the ReferIt dataset consists of 9,000 images and 54,000 referring expressions with approximately 4-6 referring expressions per image.

For the query auto-completion task, we train our extended LSTM language model where the dimension of image representation is 128, $r = 64$ is the rank of the matching personalised matrix (component from Figure 5.2). We use character embeddings with dimension 24, the dimension of the LSTM hidden units is 512, and a maximum length of 50 characters per query with Adam optimizer at a learning rate of $5e-4$ for 50,000 iterations as well as a batch size of 32. For the patch selection task, we train our proposed iBERT model using pairs of (region description, patch set) from the Visual Genome dataset, giving rise to a training set of approximately 1.73 million samples. The extra 0.3 million samples are split into test and validation sets. We conduct training for the patch selection model that fine-tunes BERT having twelve layers with a batch size of 32 for 250,000 iterations using Adam as optimiser

at a learning rate of $5e-5$ in which the performance increases steeply for the initial 10% of iterations. We use an NVIDIA Tesla T4 GPU which takes a day and a half for the complete training activity.

5.6.3 Performance Measure

We evaluate the quality of our predictions and estimations using the following performance metrics:

Mean Reciprocal Rank: The most standard metric for QAC tasks is the mean reciprocal rank (MRR), which is the average of the reciprocal ranks of the final queries in the QAC outcomes. The MRR for the query auto-completion system Q_A provided the test dataset D_T is as follows:

$$MRR(Q_A) = \frac{1}{|D_T|} \sum_{q \in D_T} RR(q, Q_A(q_p))$$

where q_p is a prefix of query q and $Q_A(q_p)$ is the list ranked for candidate completions of q_p from Q_A . RR denotes the reciprocal rank of q if q is present in $Q_A(q_p)$, in other cases reciprocal, is 0.

Language Perplexity: Perplexity is a measure to encapsulate the uncertainty of the model for a given query prefix. This metric has been explored earlier for an information retrieval task [79] and its correlation with the standard precision-recall measures has been investigated [16]. The average inverse probability is perplexity. A better model has lower perplexity.

$$Perplexity(q_p) = \sqrt[N]{\prod_{i=1}^N \frac{1}{P(q_i|q_{i-1})}}$$

where N is the normalised length of the query and $P(q_i|q_{i-1})$ is the probability of the complete query given the immediately preceding query prefix.

We evaluate the patch selection by F1 score.

5.7 Results

We report the evaluation result in Table 5.1. We perform our evaluation in two parts. Firstly, we evaluate the quality of our query completion (query prefix of length one or more characters) by mean reciprocal rank and perplexity. Secondly, we evaluate the patch selection task by F1 score. We evaluate the query completion task on Visual Genome and ReferIt datasets which have character vocabulary sizes of 89 and 77. We match index T_q of the

true query prefix in the top 10 predicted completions where we estimate the MRR score as $\sum_n \frac{1}{T_q}$ and reinstate the reciprocal rank with 0 in case the query does not appear in the top 10 completions.

Table 5.1 Evaluation results of the query completion task. Our MRR score is in bold face.

Model	MRR (Seen+Unseen)
MPC [21]	0.171
Character n-gram (n=7)	0.287
Mitra10K+MPC+λMART [145]	0.278
Mitra100K+MPC+λMART [145]	0.298
NQLM(S)+WE+MPC [159]	0.345
NQLM(L)+WE+MPC [159]	0.355
NQLM(L)+WE+MPC+λMART [159]	0.354
FactorCell [93]	0.309
E-LSTM LM(Ours)⁵	0.764

The perplexity comparison on both collections of test queries utilising corresponding contexts i.e., images and indiscriminate noise. The perplexity of the Visual Genome and ReferIt test queries with both contexts is shown in Table 5.2.

Table 5.2 Perplexity of image query auto-completion on both datasets utilising an image and indiscriminate noise. Inclusion of image results in a better (lower) perplexity

Dataset	Context	
	Image	Indiscriminate Noise
Visual Genome	2.35	3.81
ReferIt	2.63	3.45

During the evaluation of the Visual Genome and ReferIt test sets (or queries), we analyse the query prefix with different lengths for the corresponding context (noise and image). We found that mean reciprocal rank is altered by the query prefix length, as long-tailed queries are comparatively more difficult than queries of average length to match. Hence, we examine quite better performance for all prefix lengths on the ReferIt dataset (from Table 5.2).

We evaluated our proposed iBERT model for finding patch probabilities which are used to select and rank patches in the image. We instate an F1 score⁶ of 76.38% for over 3,000 patch classes.

⁵E-LSTM LM: Extended LSTM Language Model

⁶F1 score for the baseline methods shown in Table 1 were not available

5.8 Conclusion

We proposed an extended LSTM language model for query auto-completion in image search. Our modelling approach indicates that this possesses both textual information and image features in an effective manner for a significant increase in performance (in terms of MRR). The advantage of beam search for a ranked list of auto-completed queries selects an optimal auto-completion at least on a single character prefix. Our explainable approach to patch selection is backed by Information Foraging Theory, which is for the ranked patches via iBERT based on their corresponding probabilities. Overall, the auto-completed query is mapped to the most relevant image via identifying patch classes. We made an initial attempt to leverage TF-IDF in a probabilistic scenario of image patches using IFT for QAC. This theoretical explainable model is based on IFT to characterise user interaction at the query level and how the user information needs can be inferred to their goal. We intend to generalise the probabilistic patches for the interdependence of auto-completion on image selection as part of future work.

Chapter 6

Quantum-inspired Modelling for Interactive Image Retrieval

This chapter focusses on the applicability of quantum probabilistic framework for an interactive image retrieval task that presents varied challenges than image search. Interactive image retrieval enables an effective way for users to confer feedback in form of textual-visual queries. An image retrieval system typically contemplates the user's process of expressing an intent (or information need) as a query and the image retrieval's process of determining that information need. The retrieval task, in the case of interactive image retrieval, therefore, embraces an expressive user information need as a textual-visual query, where a user attempts to find an image similar to the picture composed in their mind whilst querying. In this chapter, we present *Semantic Hilbert space* (SHS), an interpretable framework that integrates the best of the two worlds - textual and image embedding approaches, respectively. The proposed framework follows the mathematical formalism of quantum probabilities to understand the relationship between a user's textual and image query inputs, given the image as an input query is considered a form of visual feedback. We discuss this quantum-inspired model for the interactive image retrieval task.

6.1 Semantic Hilbert Space for Interactive Image Retrieval

The Web searcher assumes a pivotal role in the interaction process with a search system. The primary challenge is in designing a search system that satisfies the underlying user information need. Several attempts have been made to model users' information needs effectively [154, 180, 171, 63, 105, 147, 245, 25, 121], but most prior research relied on textual and visual information representations. Additionally, users' information needs can be

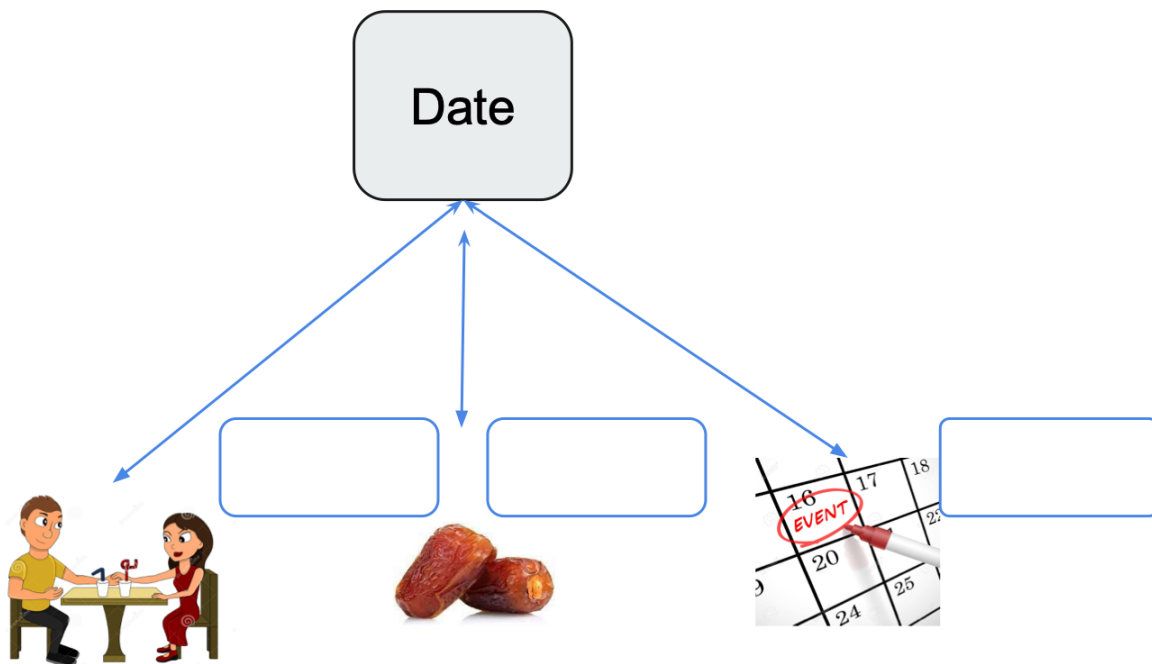


Fig. 6.1 A pictorial representation of a word ‘date’ which shows multiple meanings depending on the grounded context

either dynamic, evolving, or fixed, posing further challenges. Specifically, for image retrieval tasks, there is a need to delineate the user’s information need, as expressed by a query, in a manner that captures the user’s perception at a granular level while minimising the vagueness of representative information need.

In a standard image retrieval system, a reference or query image is used to search for a list of candidate images that match the input query through a matching process. However, such systems often fall short in supporting users in expressing their information needs as they require users to precisely convey their intentions through a single image query, thus limiting the flexibility of user interactions. In practice, it is often challenging for users to effectively communicate their perception or intention through an image query due to its inherent limitations.

Our work seeks to tackle a challenge in interactive image retrieval tasks inspired by [215]. Our task setting involves queries that incorporate both textual and visual elements, which we refer to as Intra-feedback. It is important to note that Intra-feedback is distinct from intra-query [242], as *Intra-feedback* involves incorporating grounded contextual information from a still image, which we term *guided visual feedback*. This differs from previous research [242] in the field, which has focused on considering the entirety of an image. An instance of the grounded context is shown in Figure 6.1, where the word ‘date’ depicts

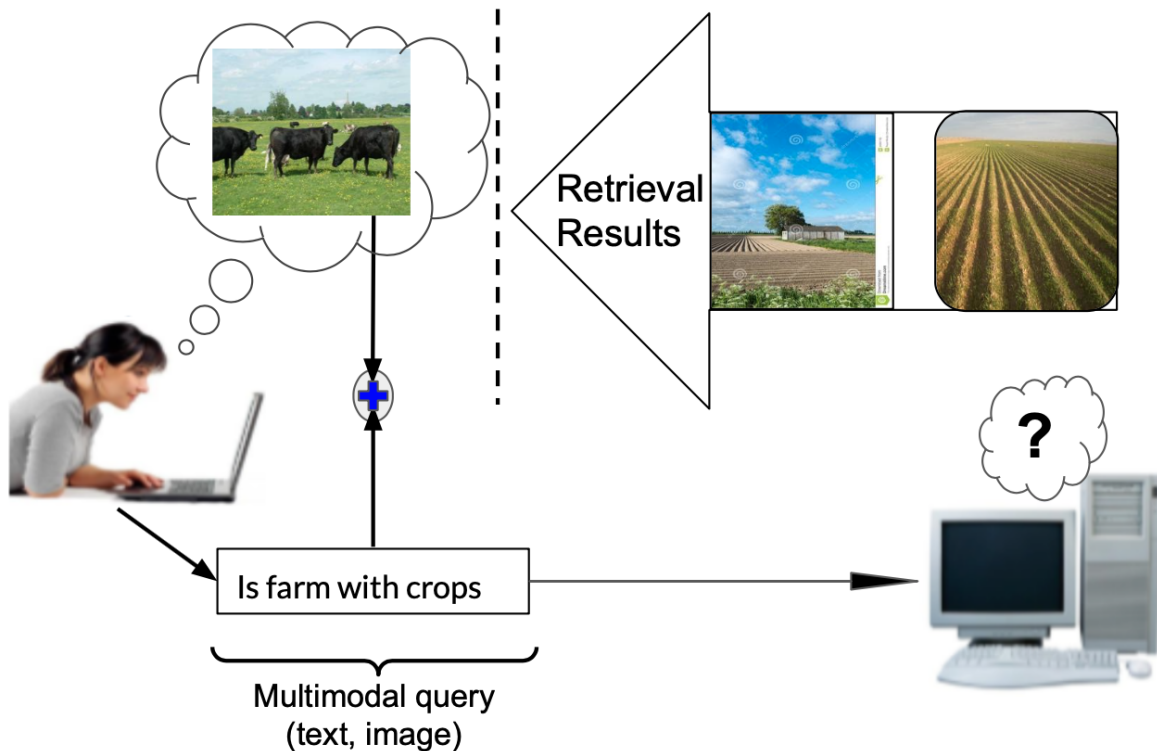


Fig. 6.2 An instance of our formulated retrieval task which depicts the expressive nature of user information needs in the context of an image retrieval scenario.

multiple meanings such as date couple (on the very left part), date fruit (in the middle), and an event date calendar (on the very right part). Conversely, a still image (the grounded context in Figure 6.1) manifests multiple words distinctively. This is also a crucial part of our task motivation, especially, the projective transformation, which in turn reconstructs image features (from the target image) back to the original input textual query. Thus, to support the concept of Intra-feedback. A mathematical interpretation of the example shown in Figure 6.1 will be illustrated in Section 6.1.2. This is integral to supporting the concept of Intra-feedback, which allows for the encoding of the cognitive aspect of a user's information need [147] using a still image and guided visual feedback, based on a semantic matching at the patch-level (where 'patch' refers to image regions). This approach leverages the Information Foraging theory [166] to provide contextual explanations through the patch selection mechanism [99]. The task overview is illustrated in Fig. 6.2, where the input information need comprises a textual query that expresses the required context for the aligned still image query, with the output being a set of retrieved results that are highly similar to the input query. The retrieval example is based on a sample query from the MIT States [92] dataset, and highlights the importance of the guided visual feedback approach in enhancing the accuracy and effectiveness of image retrieval tasks.

Drawing inspiration from recent advancements in deep complex networks [206, 124] known as complex-valued neural networks, which have been demonstrated to possess superior representational capabilities compared to their real-valued counterparts. Wang et al. [216] introduced a complex-valued DNN-driven framework that highlights the importance of phase components for words that are combined to form a sentence. Such frameworks are rooted in the existing quest [4, 2, 33] to comprehend user thought processes through the use of Hilbert space formalism [211], which underpins our work. We explore whether expressive multimodal representations of user information needs can benefit from such approaches in a more conventional setting for image retrieval tasks. In particular, we investigate whether it is possible to acquire textual-visual representations based on a complex-valued vector embedding, rather than the conventional joint vector embedding [222]. To this end, we present a complex-valued embedding strategy for learning the multimodal query. The input multimodal representation is composed of text and image features extracted using pre-trained BERT [50] and ResNet [81] models. We presume that different modality feature vectors cannot be directly combined or merged due to their distinct statistical characteristics [217] obtained from different deep neural networks. Furthermore, their representation spaces vary from joint text-image embeddings [222]. Consequently, we do not merge the input text-image query during training. Instead, our model projects image feature vectors onto textual feature vectors using an inner product, thereby learning this multimodal representation.

In order to address the challenges of interactive image retrieval, we propose a novel framework called *Semantic Hilbert space* (SHS). The SHS framework defines a common complex-valued feature space that inherently maps the input image feature (driven by Intra-feedback) to the target image feature, provided that both the input and target image features are in a common Hilbert space. We make the assumption that the input query image and target image are *Projective Transformations* of each other in a complex-valued Hilbert space, where the transformation is symmetric. This type of linear transformation allows for the integration of textual information related to the input image in the semantic Hilbert space, thereby making the user information need more expressive. We enhance the learning metric of projective transformations (PT) among different modalities features via PT symmetry, where the PT symmetry generates a representation of the input query image features from the product of the complex conjugate of the input query features with the target image features. Furthermore, the proposed semantic Hilbert space is a generalisation of classical approaches inspired by quantum probabilistic frameworks [211, 201, 171, 170], which effectively treats the cognitive and complex aspects of user's information need components (text and image query) [90, 63, 123, 216]. We evaluate the proposed SHS model on the Fashion200k [78] and MIT States [92] datasets, and compare it with several baselines. The results indicate that

our proposed method is on par with the state-of-the-art image retrieval model [215], which demonstrates the effectiveness of the SHS.

Contributions:

1. We introduce a novel interactive framework named *Semantic Hilbert space*, which aims to model multimodal information needs, specifically in the context of interactive image retrieval tasks that involve both textual and image-based queries.
2. We present a projective transformation that intrinsically maps the input image feature space to the target image feature space using complex-valued text encoding.
3. The performance of the proposed method SHS is assessed on two widely-used benchmark datasets, namely, MIT States and Fashion200k, and compared against several state-of-the-art methods. The evaluation results demonstrate that the SHS outperforms other existing methods.

6.1.1 Related Work

This section overview various existing works and methods relevant to image retrieval and multimodal representation learning. Also, we highlight the differences between these tasks and their corresponding methods.

Image Retrieval

The image retrieval process is dependent on the nature of the query input. Contemporary image retrieval systems [116] follow the ongoing trend of techniques that use deep learning. A general image retrieval system allows users to query it via a single image. This type of retrieval system belongs to content-based image retrieval (CBIR) and its application has been explored in the task of understanding user interaction using Information Foraging [129]. Recent studies in interactive image search [116] adopted attribute feedback such as ‘skateboard is blue’ empowered by effective feedback techniques but the model exhibits uncertainty in the search system. Another related area of research [121] developed an interactive lifelog search engine which allows users to query using text or images and incorporates user feedback to identify relevant images. Although user feedback can enhance the effectiveness of an image retrieval system, its impact may be constrained by the system’s indexing and feature extraction processes. This limitation can restrict the flexibility of users and hinder their ability to express their information needs, thereby impeding the system’s ability to retrieve

relevant content. To address this issue, image retrieval systems can incorporate contextual components, such as text or images, to better support users' information needs. A multimodal representation learning approach can be particularly useful in this regard, as it can enable the system to handle semantic and heuristic natural language queries more effectively, while also providing implicit and contextual information [242] that is crucial for successful retrieval.

Multimodal Representation Learning

Multimodal representation methods for image retrieval have gained popularity due to their practical application in scenarios such as product retrieval [78]. These approaches employ multimodal representation learning to create latent embeddings that associate each (text, image) pair with a dense vector that represents a point in a d -dimensional space. However, such a point fails to explicitly encode uncertainty, for instance, a word 'date' can have two or more several meanings i.e., *to eat a date*, *out on a date*, and *event date*. A few recent works have attempted to capture a word or word-meaning polarity based on complex/probabilistic word embeddings [124, 216]. Our work addresses a similar scenario to capture the contextual information from an input image so as to encode the implicit information via text but in an image retrieval task. Also, a multimodal representation has been explored in web image search [242], where the authors developed a context-aware re-ranking model for a one-to-one mapping of textual queries and image queries to a dense vector representation in an embedding space to model user preferences. Wang et al. [222] presented a joint-space embedding approach that learns both the textual and image features in a common space of identical dimension. This work demonstrates its method for two tasks i.e., image retrieval and sentence retrieval. Recently [215] developed a multimodal learning approach for image retrieval and it stands as the first work for such kind of method. However, we think their method exploits the importance of image features in different spaces and less the importance of textual features. We present these differences in the analysis part of this work. Another work addressing a cross-modal retrieval task [70] proposes FashionBERT, which performs multitask learning. This method presents a patch-level alignment i.e., image patches as tokens extracted from fashion images. A few joint representation approaches [115, 222, 1] adopted a learning mechanism of visual-semantic embedding by measuring the distance among an input linguistic query and a target image within a common feature space. However, these approaches are not suitable enough for retrieving very generic images, because users either may not have the complete picture in mind already or find it difficult to express their information need sufficiently enough via a single query.

Our work investigates how to learn a textual-visual query that can coherently express users' information needs based on the quantum probabilistic framework [211, 171], in

particular, the Hilbert space formalism. The input representation is mapped to the target image representation via the textual input which encodes the contextual information of the input still image; this encoding process is complex in nature. Therefore, we restrict the requirement of a linear transformation in contrast to affine transformation [248] due to the fact that the performance of the former transformation is superior to the latter method. We also analyse the effectiveness of the PT symmetry. The contextual encoding process from a still image follows the patch-level and text-level correspondence delineated by the Information Foraging process.

6.1.2 The Framework - Semantic Hilbert Space

We describe a framework that uniquely learns to represent the input textual feature and visual feature space, and maps the source image feature space to the target image feature space via the textual feature in a common complex-valued space, referred to as *Semantic Hilbert space* \mathcal{H} . This framework depicts representations to capture high-level ‘semantic interaction’ and serve discriminative features of multimodal components (text and image) equally in a common space i.e., Hilbert space. An overview of the framework is shown in Fig. 6.3. The

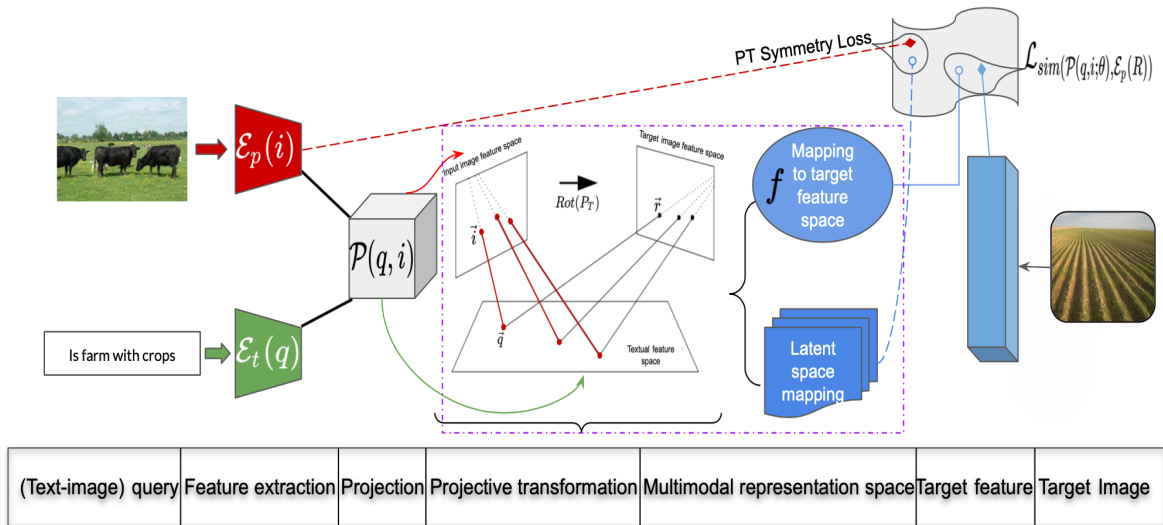


Fig. 6.3 An overview of the proposed Semantic Hilbert space framework.

input text and image query are fed into a pre-trained BERT embedding model [50] $\mathcal{E}_p(q)$ and a pre-trained ResNet34 [81] $\mathcal{E}_p(i)$ respectively. Then, a modality distribution is realised via the projection between the image features and text features. Finally, a projective transformation of the input image to the target image features in a common complex-valued Hilbert space is performed to prepare the multimodal representation, which we map in two stages. Firstly, a

mapping function (f) associates the input image features to the target feature space. Secondly, the visual guided feedback performs mapping multimodal feature space to the target feature space, and this step is referred to as latent space mapping. Then, the training loss is subjected to estimate via PT symmetry and similarity measure learning. These two mapping function inherently captures the shared multimodal representation and which leads to retrieving the closest image.

We employ the standard mathematical notations described in Section. Let a set of textual queries $\vec{Q} = \{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$, a set of image queries $\vec{I} = \{\vec{i}_1, \vec{i}_2, \dots, \vec{i}_n\}$ and a set of target image queries $\vec{R} = \{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n\}$, where \vec{q}_i , \vec{i}_i , and \vec{r}_i depicts the state vector of input textual query, input visual query, and retrieved target image. These state vectors can be treated as a *ray* in the semantic Hilbert space i.e., $|Q\rangle, |I\rangle, |R\rangle \in \mathcal{H}^1$. We use the pre-trained embedding models for feature extraction i.e., BERT embedding model [50] for textual features and ResNet-34 [81] for image features. An image pre-trained embedding model $\mathcal{E}_p(\cdot)$ extracts image features in a k -dimensional space. We use these embeddings to generate projectors in a common semantic space i.e., the projection operation $\mathcal{P}(q, i)$ of the input image $|i\rangle$ onto $|q\rangle$ is the inner product between these two and can be equated as

$$\mathcal{P}(q, i) = \langle q | i \rangle$$

Here, the projection operation can be easily interpreted from the example shown in Figure 6.1, which can be equated as:

$$|\text{Date}\rangle = a|i_1\rangle + b|i_2\rangle + c|i_3\rangle + \dots + d|?\rangle$$

where the meaning of a textual query ($|q\rangle$) ‘date’ can be manifested if its grounded context is known (the left side of the above equation). However, on the right side of above equation, it composes multiple image features (or image patches) encoded as $|i_1\rangle, |i_2\rangle, |i_3\rangle \dots$. Though, given the right side of the above equation, it reflects that combining image patches (via Intra-feedback) can retrace the original input textual query along with its context. Also, $|i\rangle$ as an image query can be decomposed into multiple image-patches ($|i_1\rangle, |i_2\rangle, |i_3\rangle$) as features in order to learn the grounded textual meaning ascertained for the given input query.

Then, the first step is to maximise the below objective function in which the projector learns the textual-visual query representation formulated as

$$\max_{\theta} \text{sim}(\mathcal{P}(q, i; \theta), \mathcal{E}_p(R)) \quad (6.1)$$

¹The state vector representations are for conciseness i.e., $|Q\rangle$ can be written as a column vector \vec{Q}

where sim is a measure of similarity and θ depicts the learnable parameter.

A better input representation (i.e., information need) enhances the learning task and in Eq. 6.1, the maximisation of the similarity measure is among the projection function output of the multimodal input query and the features of the target image. The similarity measure is not restricted to modelling only the multimodal input query but also the target image in a common Hilbert space. This way of modelling gives importance to both the text and image query. This is in contrast to [215] which defines the learning mechanism between the query image and target image and relaxes the textual features in the learning process. Our assumption to maximise the objective function (Eq. 6.1) are as follows:

- The image features of the input image query and target image are in a common space.
- The transformation (from input image query to target image) encoding uses textual features in the same common space.

Based on the following assumption, we argue that the input image query and target image are projective transformations of each other in a complex-valued Hilbert space. The intuition behind choosing such a kind of linear transformation is based on [248]. The textual features are used to compute the projective transformation. The need for complex space is for the PT to be symmetric; the textual features, when applying complex conjugate on them, can make such linear transformation symmetric. We can recall from the property of complex conjugate that the multiplication² of a complex entity (or complex number) with its complex conjugate is a real-valued number. Thus, the complex conjugate operation of a textual features vector is performed in the complex space. We arrive at

$$P_T(\vec{i}) \xrightarrow[\vec{q}]{} \vec{r} \implies P_T(\vec{r}) \xrightarrow[\overline{\vec{q}}]{} \vec{i} \quad (6.2)$$

where P_T represents the projective transformation and $\overline{\vec{q}}$ represents the complex conjugate of the textual query. Eq. 6.2 describes the projective transformation symmetry in which an input image query on such transformation from learned textual features results in the target image. Conversely, applying the complex conjugate to the textual features can transform the source image back to the original input query image. We describe the PT symmetry below in the training steps of our multimodal input representation. Also, we analyse the effectiveness of this learning metric in different settings.

²https://en.wikipedia.org/wiki/Complex_conjugate

The PT symmetry learning metric benefits from the textual feature as it helps determine the angle of such linear transformation. Such metric as a regularisation can be advantageous in terms of generalisation [150, 206].

6.1.3 Model

We have prepared the learning constructs based on different feature extraction techniques for the textual-visual query. We now incorporate them into the complex-valued embedding approach, which is an extension of [124, 216].

Feature Embedding: We generate the feature vector of an input image query via $\mathcal{E}_p(\cdot)$ in a k -dimensional space i.e., $\mathcal{E}_p(i) = i_f \in \mathbb{R}^k$. We extract text features through the BERT [50] embedding model $\mathcal{E}_t(\cdot)$ to generate text feature vectors in an l -dimensional space i.e., $\mathcal{E}_t(q) = q_f \in \mathbb{R}^l$.

Both text feature and image feature vectors are extracted from different embedding models and exhibit different statistical properties. Therefore, we perform a projection operation $\mathcal{P}(q, i)$ from an input image query to the text query. The next step is how text features encode the input image information for its transformation to the target image. To implement this, we follow our assumption in Eq. 6.2 based on the projective transformation. In particular, the projection of an input image query (features) to a target image in the complex space is done by a linear transformation. The text query features elicit the required information for the projective transformation of the input query image to the target image. This mapping of the query image and target image in a common space during training is learned via i_f and q_f in \mathbb{C}^d . Also, the angle of rotation during the projective transformation is learned via the textual features. It can be formulated as:

$$\begin{aligned} M : \mathbb{R}^d &\longrightarrow M_D \in \mathbb{R}^{d \times d} \\ Rot(P_T) &= e^{iM(q_f)} \end{aligned} \quad (6.3)$$

where M is a mapping function constructed using two fully connected layers with non-linear activation, M_D is a matrix diagonal, $Rot(P_T)$ depicts the rotation operation of projective transformation and $i = \sqrt{-1}$ depicts the imaginary number *iota*. Then, the network learns the mapping of input image features (i_f) to the complex-valued space. This mapping function is also implemented like M with fully connected layers.

$$\begin{aligned} M_I : \mathbb{R}^k &\longrightarrow \mathbb{C}^d \\ I_M &= Rot(P_T)M_I(i_f) \end{aligned} \quad (6.4)$$

where M_I is the image mapping function and I_M represents the multimodal representation of the input query.

Then, from Eq. 6.1, the goal is to maximise the measure of similarity among the input multimodal features and features of the target image. Therefore, the network requires the mapping function to learn from $f: \mathbb{C}^d \rightarrow \mathbb{R}^k$, where f is implemented using fully connected layers with non-linear activation. The mapping function associates the features from complex space to the k -dimensional real-valued space where the extracted features of the target image reside. We construct another mapping function f_l for *visual guided feedback* that is made up of two fully connected layers including a single convolutional layer, which is to capture the textual-visual similarity pattern from the data. It allows benefitting learning from local features distributed across different modalities of features. Since then the mapping functions I_M and f_l also make use of the extracted textual (q_f) and visual features (i_f) as input. The importance of visual guided feedback can be for queries that are evolving [129, 64], e.g. a user seeks a jacket with a specific logo on the left-front.

Consider a function $g(i_f, q_f)$ to delineate the input multimodal representation that learns both the textual and visual features for image retrieval. This function can be formulated as:

$$g(i_f, q_f) = \alpha f(I_M) + \beta f_l(I_M, i_f, q_f) \quad (6.5)$$

where α and β are learnable parameters.

The modelling strategy for learning multimodal representation given in Eq. 6.5 follows [201, 124], which is a complex-valued embedding network. We extend these networks for learning a multimodal representation for our image retrieval task. Having the extracted features of text and image based on the pre-trained text encoder BERT and image encoder ResNet, we use these features to learn multimodal representation in a d -dimensional complex space. A complex-valued convolutional neural network (CNN) is built to learn these representations, where the network consists of an encoder and a decoder. The complex-valued encoder network learns the multimodal representation (Eq. 6.5) and the complex-valued decoder generates the extracted textual (q_f) and visual feature (i_f) vectors back from the representation function $g(i_f, q_f)$. We adopt such a strategy to promote the image-pixel level features and to incorporate the learning of contextual information embedding within a still image. The other benefits are generalisation [150, 206] and a better representational power [216].

The complex-valued CNN network based on the above formulation contains four components which can be described as follows:

Input layer The input layer consist of a (text, image) feature vector i.e., (q_f, i_f) . Each one of the features (q_f or i_f) will be represented by a state vector.

Encoder network The encoder network for learning the multimodal representation function $g(i_f, q_f)$ (Eq. 6.5) extracts characteristic features to semantically map the input image to the most similar images (target image). For each text-image pair, the convolution can be defined based on [206], where an input is $\mathbb{I} = I + iQ$. The \mathbb{I} depicts the multimodal input matrix in which the image is realised as the real part and the word is initialised as a complex entity. The reason behind considering words as part of a complex entity is that it encodes the meaning when multiple words are combined and refer to a new meaning, and it has been used earlier in text classification tasks [124, 216]. Another reason is that encoding the context of still images and textual queries as a complex entity not only enriches the contextual information from images but also encodes such information in locating similar images for the image retrieval task. The complex-valued representation follows Eq. 6.2. Then, a matrix for the convolutional filter is $W = A + iB$, where A and B are real-valued matrices and W represents the size of a convolutional kernel. A complex-valued data matrix can be denoted as $I_f = I_{\mathbb{R}} + iI_l$, where each coefficients $(A, B, I_{\mathbb{R}}, I_l)$ are real-valued matrices. The main idea behind this encoder layer is to simulate a complex entity via real-valued entities. Now, the convolution layer can be written as

$$\begin{aligned} I_f \cdot W_m &= (I_{\mathbb{R}} + iI_l) \cdot (A + iB) \\ &= (I_{\mathbb{R}} \cdot A - I_l \cdot B) + i(I_{\mathbb{R}} \cdot B + I_l \cdot A) \end{aligned}$$

W_m depicts the matrix form of the convolution kernel, and ‘ \cdot ’ depicts the convolutional operator. Our convolutional layer is of 3×3 convolution and 64 convolutional filters. The encoder architecture is made of a fully connected layer.

Decoder network The decoder network (D_i and D_q for image and text decoder) generates back the original extracted features of text and image from the multimodal representation function (Eq. 6.5). The decoder layer first performs up-pooling and then deconvolution. The up-pooling step performs mapping the encoded features by using the location information generated from the pooling process. The second step of deconvolution re-generate the actual input text and image features from the generated sparse representation of both text and image using up-pooling.

Output The output retrieves the most closest similar image.

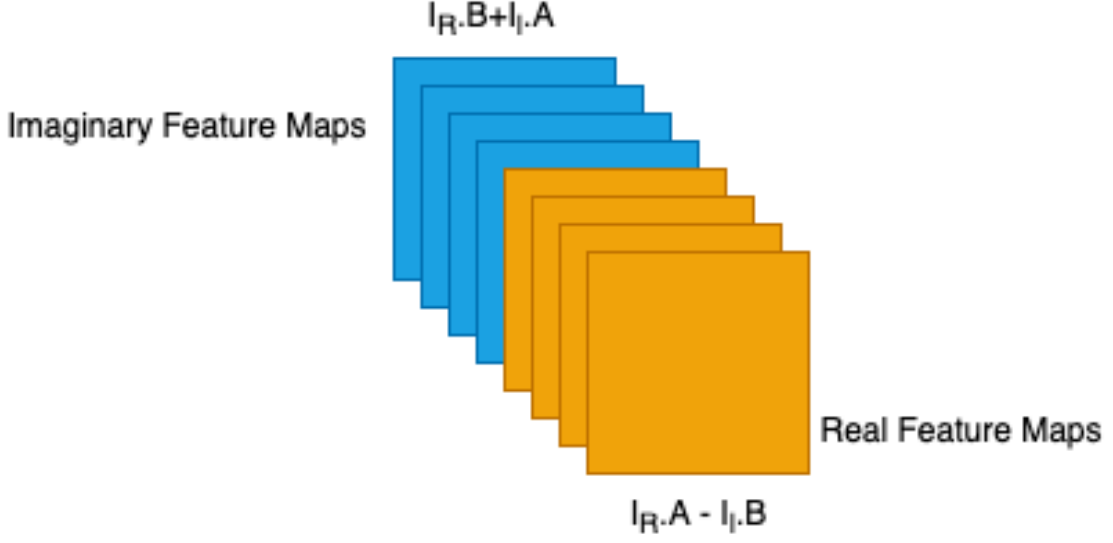


Fig. 6.4 Complex convolution layer based on Trabelsi et al. [206]

Model Training: To train our proposed model on MIT States [92], we use triplet ranking loss based on [222], which is defined as:

$$\mathcal{L}_{triplet} = \frac{1}{S \times n_{triplet}} \sum_{tr=1}^{n_{triplet}} \sum_{s=1}^S \log \left(1 + e^{\mathcal{P}(g(i_f, q_f), \mathcal{E}_p(\tilde{R}_{tr,s})) - \mathcal{P}(g(i_f, q_f), \mathcal{E}_p(R_s)))} \right) \quad (6.6)$$

where S represents the batch size for training sample s , the value in the exponent term represents a pair of multimodal features and dissimilar image feature (shown in $\mathcal{E}(\tilde{R}_{tr,s})$), and similarly, the second term with leading negative depicts a pair of multimodal features and the target image features. Each training sample s requires a fixed number of triplets $n_{triplet}$. We select the same number of triplets i.e., 3 as given in [215].

We employ the softmax loss similar to the one used in [215] for training the model on Fashion200k [78]. This loss function can be defined as:

$$\mathcal{L}_{softmax} = \frac{1}{S} \sum_{s=1}^S -\log \left[\frac{e^{\mathcal{P}(g(i_f, q_f), \mathcal{E}_p(R_s))}}{\sum_{b=1}^S e^{\mathcal{P}(g(i_f, q_f), \mathcal{E}_p(R_b))}} \right] \quad (6.7)$$

This batch classification-based loss function performs normalisation among the multimodal features ($g(i_f, q_f)$) and the features of the target image divided by the summation of similarities between $g(i_f, q_f)$ and the sample image collection in a batch (j).

The loss function for training the complex-valued CNN-based decoder is a L2 regularizer and it can be defined as \mathcal{L}_Q and \mathcal{L}_I for corresponding query reconstruction and image reconstruction

$$\begin{aligned}\mathcal{L}_{CE} &= \frac{1}{S} \sum_{s=1}^S \|q_{f_s} - \hat{q}_{f_s}\|_2 \\ \mathcal{L}_{CD} &= \frac{1}{S} \sum_{s=1}^S \|i_{f_s} - \hat{i}_{f_s}\|_2\end{aligned}\tag{6.8}$$

where $\|\cdot\|_2$ depicts L^2 norm³ and $\hat{q}_{f_s} = D_q(g(i_f, q_f))$ and $\hat{i}_{f_s} = D_i(g(i_f, q_f))$. The reconstruction loss is crafted as L^2 norm.

Projective Transformation (PT) Symmetry: We present the projective transformation of an input image query to the target image in a common space, where the angle of rotation for this transformation is estimated via textual features (Eq. 6.3) formulated in Section 6.1.3.

Similarly, for symmetry, we solicit if the product of target visual features with the complex conjugate of textual features is similar to the image query features. To do so, applying complex conjugate on the textual features in a complex space can be formulated as $\overline{Rot(P_T)} = e^{-iM(q_f)}$. Then, this joint features in the complex space form a multimodal representation for target images which are given by $\hat{I}_M = \overline{Rot(P_T)} M_I \mathcal{E}_p(R)$. This can be crafted to compute the multimodal representation formulated as

$$g(\mathcal{E}_p(R), q_f) = \alpha f(\tilde{I}_M) + \beta f_I(\tilde{I}_M, \mathcal{E}_p(R), q_f)\tag{6.9}$$

The above multimodal representation and the input image features from Eq. 6.5 are referred to formulate the PT symmetry. The PT symmetry aim to maximise the similarity function $sim(g(\mathcal{E}_p(R), q_f), i_f)$

The training loss function using the PT learning metric (\mathcal{L}_{PT}) for the Fashion200k dataset is computed by replacing the similarity function given in softmax loss in Eq. 6.7 by Eq. 6.9. It can be equated as:

$$\mathcal{L}_{PT_{Fashion200k}} = \frac{1}{S} \sum_{s=1}^S -\log \left\{ \frac{e^{\mathcal{P}(g(\mathcal{E}_p(R), q_f), i_{f_s})}}{\sum_{b=1}^S e^{\mathcal{P}(g(\mathcal{E}_p(R), q_f), i_{f_b})}} \right\}\tag{6.10}$$

Similarly, for the MIT States dataset, the training loss function follows from Eq. 6.6, where replacing the similarity function by the given equation in 6.9. The loss function can be equated as:

³[https://en.wikipedia.org/wiki/Norm_\(mathematics\)](https://en.wikipedia.org/wiki/Norm_(mathematics))

$$\mathcal{L}_{PT_{MITStates}} = \frac{1}{S \times n_{triplet}} \sum_{tr=1}^{n_{triplet}} \sum_{s=1}^S \log \left(1 + e^{\mathcal{P}(g(\mathcal{E}_p(R), q_f), \widetilde{i}_{f_{tr,s}}) - \mathcal{P}(g(\mathcal{E}_p(R), q_f), i_{f_s}))} \right) \quad (6.11)$$

The corresponding PT symmetry loss for the MIT States and Fashion200k datasets is a uniform loss function employed in training the proposed model, provided it depends on the datasets it is trained upon. We choose triplet ranking loss due to its nature of output being probabilistic, whereas softmax loss caters to identifying the output image classes, as well as can assign a probability to those classes.

6.1.4 Experiments

Dataset

We conduct experiments on the MIT States and Fashion200k datasets to validate the effectiveness of the proposed model.

MIT States [92] consists of images collected from Bing search. It contains 63,440 images representing 245 nouns altered by approximately 115 adjectives. This collection of images is human-annotated for the sake of data quality. There are certain ambiguous images and mislabelled as well. The images in the training and test set are 43,207 and 10,546. It contains 82,732 textual queries. We cleaned the dataset and removed the mislabelled and unclear images.



Fig. 6.5 An instance of sample training set images from the MIT States dataset.

Fashion200k [78] is a large collection of five different cloth categories such as dresses, jackets, pants, skirts, and tops. It contains 201,838 images. Also, this image collection is human-annotated. The training and test set contains 172,049 images and 29,789 images. The number of unique textual queries in the training set is 53,099.



Fig. 6.6 An instance of sample training set images from the Fashion200k dataset.

Evaluation

For an interactive image retrieval task, we follow the evaluation metric used in [215, 254]. We adopt *Recall@K* ($R@K$) to evaluate the retrieval performance. It is the fraction of queries for which the correct image is retrieved among the closest K points to the query. $R@K$ is the ratio of retrieved images that are relevant over the total relevant images. We perform repetition of our experiments five times in order for our performance results to be consistent. The main reason to use the Recall metric as our dataset has images containing multiple attributes or regions which inherently contribute to positive and negative images, such as the qualitative examples reported in Figure 6.7. It has multiple attributes and only the 1st, 2nd, and 4th retrieved images are relevant to the user information need. Also, the Recall metric is the number of retrieved positive images compared to the total positive images within the dataset.

Baseline Models

A range of baseline models that relates to our task closely is employed to compare our proposed SHS model. We describe these baseline models below:

- **Show and Tell:** Vinyals et al. [214] developed a deep generative model that employs training an long short-term memory network that fuses images and texts, where the image feature vector and text feature vector are directly connected.
- **Relation Network:** Santoro et al. [186] developed an augmented neural network based on relational reasoning to learn sequential object attribute features (including explicit pixel features) which mainly deals with images containing objects of the same kind.
- **Film:** Perez et al. [161] employs the Gated Recurrent Unit (GRU) as a generator and a residual block that inputs a question embedding and encodes image features. Their

key proposition is an affine transformation applied to the intermediary feature and the output of the residual neural network.

- **TIRG [215]:** It focusses on a joint-input query that comprises an input image and a text description in order to retrieve the target image. It uses a similarity measure for the relationship between the input image and the target image. One shortcoming is the ability of one-way long short-term memory to obtain features is weaker than that of models such as two-way LSTM or BERT, which affects the effect of the model to a certain extent.
- **Composed Query [86]:** This method is closely related to our task as it uses multimodal learning, where a reference image as an input with a sentence delineating the changes required for retrieving target images.

Training Setting

We adopt most of the training settings from [215] except the pre-trained image feature extractor, where we employ ResNet34 [81]. The dimension of textual feature vectors is 768 which is extracted using BERT embedding model [50]. The batch size for training is 64.

6.1.5 Results

Model \ Dataset	MIT States			Fashion200k		
	Metrics - Recall@K					
	K=1	K=5	K=10	K=1	K=10	K=50
Show and Tell [214]	11.9 \pm 0.2	31.0 \pm 0.5	42.0 \pm 0.8	12.3 \pm 1.1	40.2 \pm 1.7	61.8 \pm 0.9
Relation Network [186]	12.3 \pm 0.5	31.9 \pm 0.7	42.9 \pm 0.9	13.0 \pm 0.6	40.5 \pm 0.7	62.4 \pm 0.6
Film [161]	10.1 \pm 0.3	27.7 \pm 0.7	42.9 \pm 0.9	12.9 \pm 0.7	39.5 \pm 2.1	61.9 \pm 1.9
TIRG [215]	12.2 \pm 0.4	31.9 \pm 0.3	41.3 \pm 0.3	14.01 \pm 0.6	42.5 \pm 0.7	63.8 \pm 0.8
(+) BERT	12.6 \pm 1.0	31.6 \pm 1.0	43.1 \pm 0.3	15.2 \pm 0.4	43.4 \pm 0.2	63.8 \pm 1.2
Composed Query [86]	14.29 \pm 0.6	34.67 \pm 0.7	46.6 \pm 0.6	16.26 \pm 0.6	46.90 \pm 0.3	71.73 \pm 0.6
SHS (Ours)	14.2 \pm 0.6	36.4 \pm 0.1	48.2 \pm 0.3	23.2 \pm 0.4	55.6 \pm 1.0	74.2 \pm 0.6

Table 6.1 Performance comparison with baselines. (+) depicts the BERT model added to the TIRG for a new baseline. The best performances are in **boldface**.

We report a comparison performance of our proposed model among several existing and created baselines in Table 6.1. Our model outperforms all other methods and baselines except for the composed query [86] model at R@1 which stands best among all of the models. We also compute the test accuracy of our SHS model which is 93.44%.

Model	MIT States	Fashion200k
SHS	48.2	55.6
(+) Real space	46.0	49.2
(-) \mathcal{L}_{PT}	47.9	52.4
(-) visual guided feedback (f_l)	46.9	53.0
(-) image mapping (f)	45.4	52.6

Table 6.2 Ablation test on both MIT States and Fashion200k datasets. (+) and (-) depicts with and without the respective components to the proposed SHS mode.

Ablation Test

We analyse the effect and impact of different sub-components of the SHS model. The ablation test is reported in Table 6.2. Based on the analysis of different components, we found that the SHS model achieves significant performance among all other models. This analysis verifies that modelling both text and image query via a modality distribution realised via projection delineates the implicit contextual information of a still image as opposed to the fusing/concatenating of these two modalities. The representation power of our approach leads to effective image retrieval. We can observe in Table 6.2 that without including the loss function of projective transformation (\mathcal{L}_{PT}), the performance has a slight decrease which tells that the complex-valued representation captures the discriminant features of both modalities, and it leads to a very minor difference among the best-performed model. Another instance of the analysis is that we concatenate the real space with SHS in a complex space, and it decreases the performance significantly. To this, we can say that mapping the extracted textual and visual features into a common complex-valued space leads to a better capturing of information and encoding as opposed to the typical concatenation in real space.

Qualitative Examples

We also report a set of image retrieval instances to validate our proposed model via qualitative examples. 6.7 and 6.8 are the reported qualitative examples from both the MIT States and Fashion200k datasets. The input consists of a query image and textual query and the outputs are a set of retrieved images from the test set.

Implication

We describe a few implications of our method in order to inform if Intra-feedback (or visual guided feedback) can assist users in strengthening their information needs.



Fig. 6.7 Some qualitative examples of our proposed model. The examples are from MIT States dataset.



Fig. 6.8 Some qualitative examples of our proposed model. The examples are from Fashion200k dataset.

An instance of the multimodal query against a textual query with retrieval results is reported in Figure 6.9. This textual query and visual query are from the Fashion200k dataset



Fig. 6.9 Example retrieval from the test set based on a multimodal query against a plain textual query

In the above Figure 6.9, where a textual query with a visual query is given for the SHS model to perform retrieval. The top-right images are the original retrieval results in order for the input multimodal query. However, the bottom-right images are the retrieval results based on only a textual query.

Here, based on our ablation test reported in Table 6.2, where components are added and removed to test the effectiveness of our model. The key example shown in Figure 6.9, where only a textual query as information need is employed reflects an ablation test of the SHS model without visual guided feedback. Thus, without an Intra-feedback via a visual query (as visual information need), the retrieved images are insignificant as only the 1st, 3rd and 5th images in the bottom-right are long sleeve provided the ‘shirt’ and colour attributes are not reflective of the user information need. So, this informs that a visual query is essential for robust image retrieval (as seen in the top-right of Figure 6.9). Also, the input textual query is a complex entity based on complex-valued embedding [124, 95] which without a real entity (a visual query) is ineffective as it requires explicit feedback (visual guided feedback). Despite the fact, that when we considered a textual query as a real entity (it is reported in the ablation test Table 6.2) alongside a visual query, the retrieval performance degrades as compared to the one reported in Figure 6.9. Therefore, it is essential to incorporate a complex-valued entity that encodes the contextual meanings to reflect the fine-grained patterns (image patches) via an input visual query.

6.1.6 Conclusion

We propose an SHS framework, which could capture the implicit contextual information between an image and text (describing a multimodal information need), and effectuate the image retrieval task. The proposed model extends the classical way of representing queries and images by means of modality distribution i.e., via a projection, which leverages the projective transformation in a complex-valued Hilbert space to delineate the encoding of still images via textual features. The main idea is to realise a multimodal representation via a complex-valued CNN to enhance the image retrieval task. The experimental results on both the MIT States and Fashion200k show that our approach outperforms a number of state-of-the-art cross-modal retrieval methods, and also proves the importance of multimodal representations to represent users’ information needs.

6.2 Quantum Probabilistic Modelling for Query Interaction in Image Search

Search sessions contain multiple turns where users often refine the query when the results are unsatisfactory. A general scenario for the users’ querying could be an initial textual query that matches the description of a list of image results and subsequent multi-modal textual and image queries. Our task addresses such an image search scenario where a searcher is

attempting to find an image, or correspondingly the searcher has a generic idea of an objective image in mind or is reformulating a previous query. Initially, the searcher may begin with a textual description of the ‘belief’ [228] of the image they have in mind and subsequently explore the information space utilising a list of generic queries or topical captions or attributes (e.g., left-hand side of Fig 6.10) in the image to effectuate the search outcomes. Textual queries are supposed to describe the image but tend to be under-specified when it comes to their information needs. Image retrieval from text-based queries thus requires a certain level of semantic and visual understanding, which has gained significant improvement with the progress made in representation learning [215]. However, locating very generic images (such as on the right-hand side of Fig 6.10) with comprehensive specifications (user queries such as text, image, etc.) remains unresolved.

Current approaches [115, 222, 1] adopted a learning mechanism of visual-semantic embedding by measuring the distance among an input linguistic query and a target image within a common feature space. However, these approaches are not suitable enough for retrieving very generic images, because users either may not have the complete picture in mind already or find it difficult to express their information need sufficiently enough via a single query. Also, all these approaches fail to capture the interaction and relationship between textual and visual data. As both the linguistic and visual features delineate the common image which shares inherent correlations among them, both should be unified in the retrieval process in a complete and principled manner, which is what this work addresses.

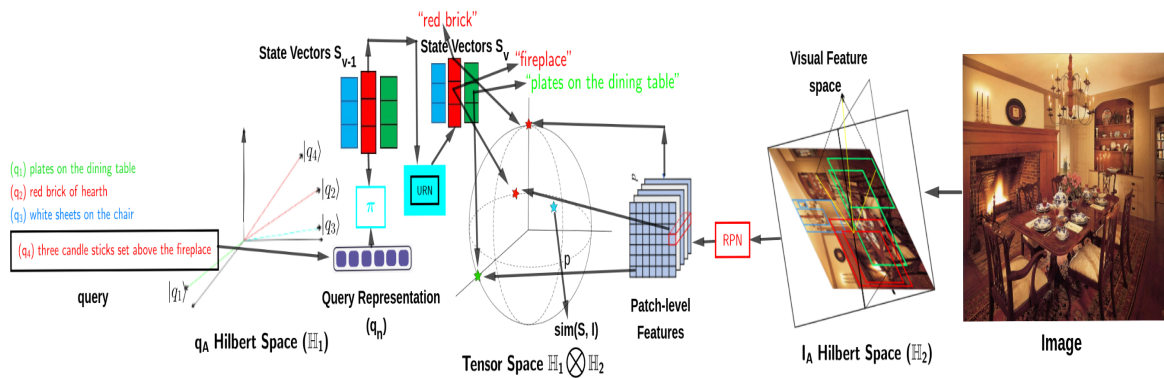


Fig. 6.10 An end-to-end architecture of Quantum-inspired Interactive model.

In Figure 6.10, the framework consists of query-assisted Hilbert space ($q_A \mathcal{H}_1$), image-assisted Hilbert space ($I_A \mathcal{H}_2$), and Tensor space (or vector space) respectively. Users’ textual queries are represented in $q_A \mathcal{H}_1$, where each eigenvector in \mathcal{H}_1 is in their corresponding sub-space as indicated by colours (or queries). The framework updates a certain set of state vectors S , modelling the historical interaction of searcher queries. For a new query q_n , the model chooses and updates one of the state vectors. The renovated state vector S_v and

image patch features are then projected to a new Tensor space (or interaction space) to measure the fine-grained alignment among each state-region duet. URN and RPN refer to the unidirectional recurrent network and region proposal network.

We present a quantum-inspired interactive framework for modelling query interactions using the Hilbert space formalism of quantum-inspired IR [211], which utilises a Tensor space⁴ (interaction space) [249] to capture the concrete alignments between multiple textual queries and images. Our motivation is two-fold:

- User information needs to be expressed as queries at each iteration may not comprehensively delineate every scope of the target image but a spotlight on certain spatial regions (delimited by information patches⁵), which allow an instinctive decomposition of the entire image within a search session [226]. Therefore, we represent images as a set of patch-level features obtained using a pre-trained object detector [174].
- Complex search sessions consist of multiple images (or patches) that might contribute to the same interaction (textual-visual feature) subspace. Specifically, neural network based model such as recurrent neural network (RNN) use their hidden states to represent sequential text queries that compress all image characteristics in a single state vector, partly able to recognise entities contributing to the same interaction subspace, such as multiple *chair* instances in the image in Fig 6.10.

To address this challenge, our framework incorporates a referring set of eigenvectors under the dynamical context of encoding textual queries corresponding to a particular image patch (or region).

The key contributions of our work can be envisaged as:

- a) Presenting an interpretable framework inspired by quantum theory for interactive image search which exploits patch-level captions as an aspect of weak supervision while training.
- b) With the usage of a large collection of linguistic-image datasets (Visual Genome [117]) to evaluate our proposed model that demonstrates the effectiveness of our model.

6.2.1 Quantum-inspired Interactive Model

We aim to model query interaction by incorporating user information needs from multi-turn search sessions (or interactive search), which retrieves images in multi-turn refinements

⁴Tensor space is to capture the non-separability of linguistic and visual features

⁵*Information patch* as known from Information Foraging theory

which extenuate the vagueness of each query. We hypothesise that searchers are commonly undetermined in their queries, which can be mitigated by the system to help them signalling to spatial patches (or regions) of the target image. We thus learn these patch-level alignments in a tensor space by capturing to map L textual queries q_l with $1 \leq l \leq L$ and the target image I_t in Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , respectively, into two sets of eigenvectors $|u_i\rangle$ ($1 \leq i \leq N$) and $|v_j\rangle$ ($1 \leq j \leq M$). We then introduce how to score the matching of q_l and I_t in the tensor space using quantum-like measurement, and collecting fine-grained symmetry among the $|u_i\rangle$ and $|v_j\rangle$. The quantum-inspired framework for query interaction is depicted in Figure 6.10.

Query Representation in Hilbert Space

We regard queries as observables as known from quantum-theoretic systems, where a system state can be computed by means of certain physical operations and is observable to it [170]. For multi-turn image retrieval, there needs to be a state representation that incorporates queries from multi-turn search sessions. Recent work [200, 49] adopted variants of recurrent neural networks that model historical context and a single query in distinct neural network modules. These methods generate a single state vector that blends all queries. However, quantum probabilistic [64] and quantum language model [123] approaches have been adopted to model a query in multi-query session search but are limited to document retrieval tasks. The key challenge our model focusses on is to represent queries to retrieve images that contain multiple objects, in contrast to current state-of-the-art models [76] using single state vector representations and which are less effective for images having multiple objects. Therefore, these approaches use image features commonly extracted from the final layer of pre-trained object detection or image classification model, and input components of the very identical or same classes enable the common feature units in the generated feature space. Hence, it is significant for these observable representations to encode and differentiate multiple patches from the very identical or same classes (such as multiple *chair* instances in the image in Fig.6.10).

We present to update a set of eigenvector representations $S = \{|u_i\rangle | i = 1, \dots, N\}$, $u_i \in \mathbb{R}^D$ for multi-turn queries. Here N^6 denotes the number of eigenvectors. We represent queries as observables in Hilbert space (\mathcal{H}_1) in the left part of Figure 6.10. Based on the quantum probabilistic framework, we measure eigenvalues of the observable to know in which state it collapsed. This complies with the notion of quantum measurement discussed in theoretic models of IR [211, 64], where a query as observable can be measured on a visual feature space by the similarity between the textual feature (i.e., term) and visual feature of an

⁶ N delineates the computational cost as retrieval time depends on the representation compactness

image. The higher similarity value depicts the higher relevance of the image (or patch) to the query. The density of a query is $\rho_q = \sum_{i=1}^n q_i^2 |f_i\rangle \langle f_i|$, with f_i the visual feature of the i th dimension in the feature space. To represent the observable, we use a spectrum for query $q = \sum_{i=1}^n \lambda_i |e_i\rangle \langle e_i|$ where λ_i is the eigenvalue and $|e_i\rangle \langle e_i|$ the projector to the corresponding eigenbasis.

Users might confer a generic description of the image in the initial turn of querying, with succeeding queries delineating distinct patches. The main goal is to investigate a better alignment among queries and image patch representations $|v_j\rangle$. An optimal set of $|u_i\rangle$ should encode and learn to assemble the user input queries into visually dissimilar representations signalling specific image patches. To encode a query, we assume that each input query delineates a patch (region) of the image and each q_t (query at time t) updates a single state vector $u_m^{t-1} \in S^{t-1}$. In Figure 6.10, the model maps each term token in the query to an \mathbb{R}^n dimensional vector space through a linear projection, then uses a unidirectional recurrent network (URN) i.e., gated recurrent units (GRU) to create the sentence embedding. Thus, the probability of a query term (u_m^{t-1}) sampled from the last immediate state vector set S^{t-1} is

$$\pi(u_m^{t-1} | S^{t-1}, q_t) = \frac{e^{f(u_m^{t-1}, q_t)}}{\sum_j^M e^{f(u_j^{t-1}, q_t)}} \quad (6.12)$$

where $f(u_m^{t-1}, q_t) = W_\pi^3(\delta(W_\pi^2(\delta(W_\pi^1[u_m^{t-1}; q_t] + b_\pi^1)) + b_\pi^2)) + b_\pi^3$ is a multi-layer perceptron which maps the connection of u_m^{t-1} and q_t into a scalar value. The probability value in equation (6.12) gives 1 if S^{t-1} is an empty vector. The model parameters are $W_\pi^1 \in \mathbb{R}^{D \times 2D}$, $W_\pi^2 \in \mathbb{R}^{D \times D}$, $W_\pi^3 \in \mathbb{R}^{1 \times D}$, $\{b_\pi^1, b_\pi^2\} \in \mathbb{R}^D$, $b_\pi^3 \in \mathbb{R}$ and δ denotes the ReLU activation function. We initially update q_t to a void state vector and after sampling u_m^{t-1} , we update this state vector via GRU. The learnable parameter for updating the state vector is $\pi(\cdot)$ which includes all the six model parameters mentioned before in this section. In the next subsection, we describe the method to update the state representations $|u_i\rangle \forall i = 1, \dots, N$ from the queries q_l , $l = 1, \dots, L$ in order to improve their score of matching with the target image.

Image Representation in Hilbert Space

We represent images by their entire visual features altogether in a new Hilbert space \mathcal{H}_2 reported in Figure 6.10. We use [9] to detect candidate patches envisaged in queries. To identify patches in an image, we utilise a region proposal network (RPN) [174] (i.e., Faster-RCNN). Also, this object detector region-of-interest (ROI) pooling layer extracts associated features a_f with the image patches. In this work, we exploit the object detector [9] pre-trained on Visual Genome [117] with 1600 predetermined image patch classes. To extenuate a_f

into a set of D -dimensional eigenvectors $V = \{|v_j\rangle | j = 1, \dots, M, v_j \in \mathbb{R}^D\}$, we apply a linear projection $v_j = W_I a_f + r_I$. Here M refers to the number of patches in each image and r_I is the bias (or randomness) in the image. For convenience, we represent the learnable criterion by $\theta_I = \{W_I, r_I\}$.

We compute the similarity score between S and V , a query term, and an image patch, formed on the term's correlations with each visual feature dimension. Each potential state-patch pair $(u_i, v_j) : sim(u_i, v_j) = u_i^T v_j / \|u_i\| \|v_j\|$, where $\|\cdot\|$ represents the L^2 norm. We use this similarity score to measure the similarity $sim(u_i, I)$ between a state vector (u_i) and the target image I as

$$sim(u_i, I) = \frac{1}{M} \sum_{m=1}^M c_{im} sim(u_i, v_m), c_{im} = \frac{e^{sim(u_i, v_m)/\sigma_h}}{\sum_j e^{sim(u_i, v_j)/\sigma_h}} \quad (6.13)$$

where σ_h is a temperature hyperparameter. We adapt an identical formulation from [134] to measure the similarity between u_i and a context vector $\sum_{m=1}^M c_{im}$. Therefore, the similarity value for the joint representation is $sim(S, I) = \frac{1}{N} \sum_{m=1}^N sim(u_m, I)$.

6.2.2 Experiments

Evaluations are conducted on the Visual Genome dataset [117]. Every image in this dataset is annotated by multiple region captions. The data preprocessing is to remove doublet region captions i.e., multiple captions that are completely identical, and images that contain less than ten region captions. The preprocessed set contains 105,414 images of which were taken 92,105 for training, 9,896 for testing, and 5,000 for the validation set. Also, we ascertain that the test set images are not employed for the training of the object detector [9]. We consider region-wise captions as queries for the training of our model. The vocabulary size (14,000) of queries is generated from the terms that embark more than 10 times in entire region captions. We randomly order queries during the training step, whereas in the testing/validation step, the orders of the queries are determined. We carry out detailed experiments to verify the effectiveness of the proposed model (QIIM), and compare the following models (Table 6.3)

- **HRED**, the hierarchical recurrent encoder-decoder network [200, 76] which is reckoned as the baseline model
- **P-HRED**, a patch-oriented HRED model derived from baseline (i) which is trained using the image patch features $|v_j\rangle$
- **P-RE**, a patch-oriented model identical to baseline (ii) but it uses a unidirectional GRU network (to encode the queries concatenation) instead of a hierarchical text encoder

- **QIIM**, the proposed quantum-inspired interactive model. Here, $\text{QIIM}_{N \times D}$ represents the model with N state vectors, each having the same dimension of D .

Model	Query Dimensions	Image Size	Parameters
HRED	1080	1080	22,820k
P-HRED	512	36×512	9,866k
P-RE	1080	36×1080	22,820k
$\text{QIIM}_{5 \times 256}^{**}$	1280	36×256	5,830k
$\text{QIIM}_{3 \times 128}^*$	384	36×128	4,861k
$\text{QIIM}_{3 \times 256}^*$	768	36×256	5,830k

Table 6.3 Specifications of the Image and Query Representations

A number of parameters are embraced in the proposed model. For convenience, we train all the models with ten-turns queries to demonstrate the effectiveness of baselines and the proposed model in both short-term and long-term queries. We extract the top 36 patches for each image using a pre-trained FasterRCNN model [9]. In the entire experiment, we set the hyperparameter (σ_h) value to 9 and Adam for optimisation with an initial learning rate of $2 \cdot 10^{-5}$, and batch size of 64. The proposed model and baseline models are validated through each epoch and trained on over 250 epochs. We used the best-performed model on the validation set for evaluation.

6.2.3 Results

We conduct an evaluation of the cross-modal retrieval task to verify our model approach. For 10 query sessions, we evaluate the retrieval performance in terms of Recall@K (R@K) i.e., the fraction of queries for which the correct image is retrieved among the closest K points to the query (see Table 6.4).

Model	R@1	R@5	R@10
HRED	32.1	60.2	70.3
P-HRED	38.1	65.0	78.3
P-RE	35.0	64.3	76.4
$\text{QIIM}_{5 \times 256}^{**}$	59.6	81.2	88.2
$\text{QIIM}_{3 \times 128}^*$	48.0	73.1	81.3
$\text{QIIM}_{3 \times 256}^*$	49.8	74.1	82.0

Table 6.4 Performance on Visual Genome where query representations is of same (*) and varied memory size(**)

6.2.4 Conclusion

We presented a novel quantum theory-based interactive model for query interaction in a multi-query image retrieval task (session search). The proposed model extends the classical way of representing queries and images by means of Hilbert space formalism, which leverages the quantum probability to delineate the joint dynamics of users' information needs evolving in historical interaction. Our proposed QIIM approach outperformed the classical baselines by a margin and its potential is likely to benefit from further enhanced interactive search. Also, it addresses the challenges of multi-turn image retrieval such as query state representations and patch-level features. The tensor space confers a formal way to intact cross-modal feature spaces for being both modalities entangled and possibly enable interactive search in a principled framework. We also found that the patch-level features play a significant role in retrieving images for a specified query, which is due to the patch-aware alignment (relevant/irrelevant signals which means patches or objects that overlap or disjoint such as Fig. 2 in [9]) matches one of the topical captions from the collection.

Chapter 7

Conclusions

7.1 Summary of Work

The basis of this thesis is rooted in quantum theory and in particular, the Hilbert space formalism for IR can be employed to model the user's adaptive information needs more emphatically than extant methods.

The focus is to develop a new representational framework that can elicit the user's dynamic information needs which are vague and evolving in nature. We deliberated the emphatic modelling approach, the quantum probabilistic framework (Chapter 2). The level of constructs in such a framework is adaptive and interactive in nature which stands as a promising method for our modelling problem. The interactive nature of the constructs is strengthened with the usage of the Information Foraging Theory. From the perspective of IFT, the searcher aspect is formulated based on the constructs of IFT.

We subsequently renew our focus to examine how quantum probability theory can be utilised to enhance word embeddings for the searcher behaviour (Chapter 1 - RQ 1). This task employed the generalised approach of word embedding using a quantum probabilistic framework to frame a classification task in question answering, which is analogous to the problem of identifying relevant documents in a collection. It is reported in Chapter 3. This gives rise to a quantum-inspired word embedding which was used to express the interactive nature of a searcher. We then employed the notion of reinforcement learning and proposed a theoretic quantum-inspired reinforcement learning framework to guide the information seeker (or a searcher) in an unknown search space. The generalised word embedding is a key actor to designate a reward value (or relevance) to the user's information need (or action) as explicated by a textual query. The parameterisation of our reinforcement learning framework employs the constructs of quantum probability and partially from the quantum language model. In a policy network, the Actor-Critic method was used, where the Critic

network is controlled by the quantum-inspired word embedding model. This led to believe that the classical model generalises using the framework of quantum theory and is consistent with the notion of quantum probability as opposed to classical Kolmogorovian probability theory [176]. That is, the quantum probabilistic framework emerges to be more competent to model the information seeker (searcher or forager) behaviour.

The searcher behaviours are not reliant on a single aspect and so the role of IFT stands as a key part in knowing what influences their behaviour either in the positive or negative sense. It means that certain aspects can enhance the searcher's behaviour and some of them can be of partially meaningful impact. To do this, we facilitate the task of how Information Foraging Theory can explicate and enrich the user interaction mechanism in IR. We first examine the user-item interaction process in a recommendation task where there is no involvement of query, however, searcher preferences became the most important aspect in assisting the interaction process among the recommended items and user selection (RQ 2). This is comprehensively detailed in a content-based image recommendation task which is reported in Chapter 4.

We also considered an IR task i.e. query auto-completion for image search, which involves reformulation and expansion of query simultaneously, and also exhibits the level of user's information needs by means of information scent levels and patterns (Chapter 5 - RQ 2). We defined the notion of user interaction in search completion, where each typed prefix by a user generates a set of auto-completed query which resembles their information needs. Later, the user information needs to elicit whether the retrieved image result reflects their aligned region of interest (or patterns of information goal delineated by information scent). We unveiled that information scent determinately explicates the user information needs which continues to be vague and evolving based on the patterns and levels of information scent.

This unfolds the intuition of integrating such a behavioural model - IFT within the quantum probabilistic framework to whether it enhances the user information needs, and unifies under a principled framework that can characterise the user interaction mechanism (Chapter 6 - RQ 3). The very first motive to the interactive search is rooted in the visual information seeking [73] which describes that an image as a query can better explicit their information needs even if the textual queries are shorter. To facilitate this, we proposed a quantum-inspired interactive framework that models the user's multimodal information needs. The user's information needs comprises a textual query and a visual query of which both are neither concatenated nor joined, rather a projection is used among it to transform the visual query to the target image query via a projective transformation. The input visual query act as guided feedback based on IFT that can harmonise textual query in a way that helps transform to the target image. A projective transformation approach is presented which

employed the complex conjugate operation and so the nature of transformation is complex (in complex-valued space). The entire process of transformation is under Hilbert space that composes the feature encoding of each modality. The framework is referred to as semantic Hilbert space, where the ‘semantic’ is named due to the multi-semantic nature of the user information needs, for interactive image retrieval tasks. We further explored a traditional way of treating the user information need - a textual query. However, in a session search, where the information needs is a set of multi-turn queries. We model such information needs using the Hilbert space formalism, in an image retrieval task. The representation of both modalities is explicated in their own Hilbert space. So, here, the similarity between input textual query and target image employs Tensor space operation for integrating the different modality features to be entangled. The set of a user’s information needs is modelled for query interaction and thus, to enrich the image retrieval effectiveness.

In particular, we witnessed improvements in our proposed model based on the quantum probabilistic framework, and comparatively better performances than those of existing classical approaches, such as cross-modal retrieval.

7.2 Main Findings and Limitations

The premise of the work outlined in this thesis is to model the user information needs in an interactive IR task based on Information Foraging Theory. A broad research question has been formulated in Chapter 1, which comprises three sub-research questions. I highlight the main findings and limitations of the work carried out with respect to the following three research questions:

RQ 1: How can a quantum probability view of word embedding be applied in information retrieval, and in particular to information seekers?

Main Findings: This research question is comprehensively answered in Chapter 3 and reported the contribution of it in Section 1.3.1. There are two sections described within this Chapter, where the first experiment is to explore and investigate the quantum-inspired word embedding for textual classification, in particular, question answering. The probabilistic (quantum probability) viewpoint to classical word embedding is posed due to the implicit nature of words when combined to form a new meaning or multiple meanings. This revisits the need for quantum-like word embedding. The second section in Chapter 3 utilises the aforementioned quantum-like word embedding in a classical reinforcement learning framework to learn a searcher behaviour in an unknown search environment. This part of the

work is tailored to devise a theoretic framework of quantum-inspired reinforcement learning, which first formulates the information seeking task as reinforcement learning with the usage of Information Foraging Theory. This process of translating the information seeking task is termed ‘Reinforced Foraging’. We found that there is more to IFT when integrated into an RL mechanism. Then, it is encapsulated within a reinforcement framework parameterised via quantum probability constructs in a query matching task. The entire framework constitutes two parallel representations - the user actions (information needs) and their corresponding action states. The modelling approach is inspired by Hilbert space formalism which is the backbone of quantum probabilistic framework and can be scaled/generalised to varied IR tasks.

Limitations: The first part of the work which investigates quantum-like word embedding can be further evaluated either on varied datasets in a word-/sentence-related task or on the merits of the trained complex-valued embedding vectors. The complex-valued word representation can be reasoned as to whether it depicts a true representation of a word/sentence. During the training of such complex embedding vectors, the duplicates or similar vectors are pruned. This can be better with respect to those datasets which has fewer contextual words/sentences, however, if it happens to be in a higher ratio, then these complex embedding vectors be short of finer information if compared with classical probabilistic word embeddings. In the second work, we proposed a theoretic quantum-inspired RL (qRL) framework to learn and guide the searcher’s behaviour. However, it can be used to evaluate with certain datasets for query matching or summarisation tasks. This qRL framework can be employed in a different context to generate users’ behaviour data if given a set of requisite parameters, to simulate the search environment.

RQ 2: To what degree can user interaction mechanism be explained by Information Foraging Theory?

Main Findings: The applicability of IFT in some of the IR and recommender system tasks is presented in Chapter 4 and Chapter 5. The contribution of it is reported in Section 1.3.2 and Section 1.3.3 referred in Chapter 1. Firstly, I formulate IFT-based strategies in a content-based image recommender system to enhance the user information needs by personalising their preferences and finding implicit behavioural signals that user follows to select images on the recommendation. This first experiment exploits the Pinterest image collection and WikiArt dataset. We describe two sub-tasks of which, the first sub-task is to understand the impact of user attention on an image recommendation. The second sub-task is

to incorporate implicit behavioural features in an image recommendation. We found that the recommendation enhances if the information scent of the selected item (image) is strong. The information scent of an image magnifies with the use of visual cues (or visual bookmarks). The other part is to understand how information scent influences search preferences. In this, we found that textures within an image assumed as a cue exhibit strong information scent, and emerge to be better if such cues are in images that have objects embedded within it of large size. This also tells us the improvement in the user perception whilst they pick from one of these recommended images. Also, the search preferences are effective due to ranking of relevant images for any selected preference elicits the top items which have visual or textual cues in them.

The second task (Chapter 5) aims at understanding user interaction at the query level in the real-time image query auto-completion task, and how the user information needs to explicate searcher attention based on IFT. In this experiment, we formulate a task of query auto-completion for image search, where a prefix typed by the user is auto-completed using our extended LSTM language model, and then subjected to an image to reflect the relevance (or attention) of the predicted query. The auto-completion language model utilises beam search to find the optimal query for the input prefix. Then, we make use of the BERT model to extend for identifying the image patch classes and ranking these patches in order to reflect the auto-completed query. Also, we propose an IFT-based strategy to explain the tendency of image patches to explicate the user information needs.

These two tasks where IFT plays a key role in explaining the user behaviours and aspects related to the information needs of a searcher. This embarks the importance of behavioural models integrated into a principled framework that enriches user interaction.

Limitations: The evaluation aspect of content-based image recommendation that follows the IFT-based strategy uses a small set of image collections, which is one of the limitations. It can be reasoned as to what extent these experimental scenarios depict real-time image recommendation cases. The second task of image query auto-completion lacks user evaluation. We used a large collection of images and textual queries dataset that can be used to reflect real-world scenarios if the task is catered by users. That can also distinguish the interpretation of results obtained from the user and system evaluation.

RQ 3: Whether the user interaction mechanisms using Information Foraging Theory could inform effective formal (quantum) models for interactive search?

Main Findings: In answering this research question in Chapter 6, we find that our proposed

method based on Hilbert space formalism can effectively model the user's multimodal information needs and perform better in terms of Recall score. My contribution to it is reported in Section 1.3.4. Firstly, we formulate multi-semantic information needs that are expressive to be explicated by the user, to capture the implicit contextual information among their query and an instance of the image as visual feedback. The representation of this multi-semantic information needs to utilise Hilbert space and we propose a projective transformation to distribute the modality using textual features in a complex-valued space. A major impact of this finding is that it generalises a classical retrieval task and effectuates the usage of the quantum probabilistic framework. Secondly, we also investigate a task of multi-turn queries (evolving information needs) in an image retrieval scenario. We propose a quantum-inspired interactive model to represent textual queries and images by means of Hilbert space formalism. The user information needs to explicate the historical interaction by means of representation in a quantum probabilistic framework. We find that our approach performs better against comparative classical neural models. This also reports empirical results in Chapter 6 and reflects the applicabilities of quantum theory to image retrieval-related problems, as introduced by [211, 232], can lead to advancements in image retrieval. This notion is in affinity to the contention solicited by [170].

Limitations: The experimental settings employed for the first task of modelling multimodal information needs can be enhanced if evaluated on varied datasets such as Artwork related images and queries. We refer to a specific dataset due to the fact that our model has utilised datasets which has multiple objects (or patches) in an image. However, it can be argued whether our model is effective against a dataset that has textures within patches in an image (such as Artwork images). The second proposed model - the quantum-inspired interactive model can have one limitation of baseline model selection. We have instead derived a set of baseline models and compared the proposed model against it in Chapter 6.

7.3 Future Work

This work conferred in this thesis provides highlights and approaches for explicating the user information needs in IR based on IFT, including insights into user behavioural aspects. Apart from the aforementioned vital findings and summary, it opens up several prominent directions for the scope of subsequent work. We elucidate possible follow-up research direction with respect to the formulated research questions solicited in this thesis.

7.3.1 User Interaction in Image Query Auto-Completion

A general scenario for the users' in a query auto-completion system could be an initial text prefix that matches one of the search completion from the suggestions list and subsequent alignment to the image result reflects their information goal. A web searcher is a primary actor in the process of interaction with a search system. A specific challenge lies in how the search system caters to the underlying user information needs. We discuss the aspect of such a scenario in our work reported in Chapter 5. However, the consequential alignment of an auto-completed textual query to images captures the user's perception but only upon a condition where the input to this alignment model is a textual query that has pre-trained image features. There is a scope to renovate the input query by extending it to multimodal (textual-visual query). The need for such an enhancement is required to delineate the implicit contextual information based on their vague prefixes. For instance, the prefix 'Asian' or 'Asian restaura' is invoked by the user in a query auto-completion (or search engine) system, provided he/she expects to receive suggestions about an Asian restaurant around their local place. However, the word 'Asian' is contextual and search completions may be biased toward popularity instead of satisfying the user's vague information need. However, the user information need can be informative if the search completion system asks him/her to provide an instance of the exemplary image. This visual information can be user-centric based on the provided image in the input, and that can be meaningful to the user for finding their favourite restaurant instantly. The application of such a search completion system is what real-world scenarios expect to provide and integrate within existing search engines. At a broad level, its impact of it can be applicable to domain-specific searches such as grocery search, restaurant search, medicine search¹, car search, classified advertisements search² etc.

7.3.2 Incorporating User Behavioural Features in IR System

Generally, the text search systems follow historical click logs, search logs, user behaviours (implicit and explicit), and user-item interaction datasets to explicate the user's information needs. However, these search systems explicitly lack semantic information and fold meaningful results. We have carried out an extensive investigation in this area, especially the inclusion of semantic information in users' information needs. Also, in Chapter 6, we tackled most of them in an image retrieval task. However, our proposed models are specialised and not tailored to some real-world scenarios. Those real-world scenarios are local search, social search, etc. involving geographic information and numerous facets of user behaviour. One

¹<https://pharomeasy.in>

²<https://www.avito.ru/>

probable reason for the applicability of our model in such scenarios is due to the intuition behind users, as they tend to consider acts that are closer to them. This reason can be analogous to the previous sub-section 7.3.1 where users prefer daily activities such as grocery search, restaurant search, etc. The challenge of an image retrieval task is to delineate a user's information need (as expressed by the query) in a way that captures the user's perception at a granular level. However, if considered a local search engine that required local queries that reflect geographical information, this kind of searcher also requires the user behaviours such as click signals, geographical information (distance, location, etc.), the popularity of items area-wise and so the quality of the items or service, etc. Alternatively, to capture the interdependence between different user behaviours using specific sensors, virtual reality devices and internet-of-things (IoT) can be helpful to enrich or infer the information to the user via mobile devices. Mobile devices are widely used to keep a record of interactions with different applications, for instance, web browsers act as a medium to capture such user-related data. However, it can be amplified if sensors incorporate it for data collection of such granular local searches. These are some very relevant real-world scenarios for which our work at such a broader level can be applied and potential to scale in different domain-specific searches.

References

- [1] Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., and Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, 164:116–143.
- [2] Aerts, D. (2014). Quantum theory and human perception of the macro-world. *Frontiers in Psychology*, 5:554.
- [3] Aerts, D., Arguëlles, J. A., Beltran, L., Beltran, L., Distrito, I., de Bianchi, M. S., Sozzo, S., and Veloz, T. (2018). Towards a quantum world wide web. *Theoretical Computer Science*, 752:116–131.
- [4] Aerts, D., Gabora, L., and Sozzo, S. (2013). Concepts and their dynamics: A quantum-theoretic modeling of human thought. *Topics in Cognitive Science*, 5(4):737–772.
- [5] Agichtein, E., Brill, E., and Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 19–26. ACM.
- [6] Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3):261–295.
- [7] Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, 98(3):409.
- [8] Anderson, J. R. (2013). *The adaptive character of thought*. Psychology Press.
- [9] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., and Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6077–6086.
- [10] Andrilli, S. and Hecker, D. (2010). *Chapter 7 - Complex Vector Spaces and General Inner Products*. Academic Press, Boston, fourth edition edition.
- [11] Arafat, S., Van Rijsbergen, C., and Jose, J. (2005). Formalising evaluation in information retrieval.
- [12] Arafat, S. and van Rijsbergen, C. J. (2007). Quantum theory and the nature of search. In *AAAI Spring Symposium: Quantum Interaction*, pages 114–121.

- [13] Asano, M., Basieva, I., Khrennikov, A., Ohya, M., and Yamato, I. (2013). Non-kolmogorovian approach to the context-dependent systems breaking the classical probability law. *Foundations of physics*, 43(7):895–911.
- [14] Azzopardi, L. (2014). Modelling interaction with economic models of search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 3–12. ACM.
- [15] Azzopardi, L. (2017). Building cost-benefit models of information interactions. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 425–428. ACM.
- [16] Azzopardi, L., Girolami, M., and Van Rijsbergen, K. (2003). Investigating the relationship between language model perplexity and IR precision-recall measures.
- [17] Azzopardi, L. and Zuccon, G. (2016). An analysis of the cost and benefit of search interactions. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, pages 59–68. ACM.
- [18] Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.
- [19] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- [20] Balabanović, M. (1998). Exploring versus exploiting when learning user models for text recommendation. *User Modeling and User-Adapted Interaction*, 8(1-2):71–102.
- [21] Bar-Yossef, Z. and Kraus, N. (2011). Context-sensitive query auto-completion. In *Proceedings of the 20th international conference on World wide web*, pages 107–116. ACM.
- [22] Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982). Ask for information retrieval: Part i. background and theory. *Journal of documentation*.
- [23] Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155.
- [24] Berberich, K., Bedathur, S., Alonso, O., and Weikum, G. (2010). A language modeling approach for temporal information needs. In *European conference on information retrieval*, pages 13–25. Springer.
- [25] Berget, G. (2020). " information needs of the end users have never been discussed" an investigation of the user-intermediary interaction of people with intellectual impairments. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 93–102.
- [26] Berget, G. and Sandnes, F. E. (2019). Why textual search interfaces fail: a study of cognitive skills needed to construct successful queries. *Information Research: An International Electronic Journal*, 24(1):n1.
- [27] Berry, M. W. (2001). *Computational information retrieval*, volume 106. SIAM.

- [28] Blacoe, W. (2015). On quantum generalizations of information-theoretic measures and their contribution to distributional semantics. *arXiv preprint arXiv:1506.00578*.
- [29] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- [30] Brennan, K., Kelly, D., and Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 165–174. ACM.
- [31] Brill, E., Lin, J. J., Banko, M., Dumais, S. T., Ng, A. Y., et al. (2001). Data-intensive question answering. In *TREC*, volume 56, page 90.
- [32] Bruza, P., Kitto, K., Nelson, D., and McEvoy, C. (2009). Is there something quantum-like about the human mental lexicon? *Journal of Mathematical Psychology*, 53(5):362–377.
- [33] Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.
- [34] Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., and Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American psychologist*, 53(5):533.
- [35] Cai, F., De Rijke, M., et al. (2016). A survey of query auto completion in information retrieval. *Foundations and Trends® in Information Retrieval*, 10(4):273–363.
- [36] Campbell, I. and van Rijsbergen, C. J. (1996). The ostensive model of developing information needs. In *Proceedings of the 3rd international conference on conceptions of library and information science*, pages 251–268. Citeseer.
- [37] Cao, H., Jiang, D., Pei, J., He, Q., Liao, Z., Chen, E., and Li, H. (2008). Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 875–883. ACM.
- [38] Card, S. K., Pirolli, P., Van Der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., and Boshart, J. (2001). Information scent as a driver of web behavior graphs: results of a protocol analysis method for web usability. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 498–505. ACM.
- [39] Chandak, Y., Theodorou, G., Kostas, J., Jordan, S., and Thomas, P. S. (2019). Learning action representations for reinforcement learning. *arXiv preprint arXiv:1902.00183*.
- [40] Charnov, E. L. et al. (1976). Optimal foraging, the marginal value theorem.
- [41] Chen, J., Zhang, H., He, X., Nie, L., Liu, W., and Chua, T.-S. (2017). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 335–344. ACM.

- [42] Chi, E. H., Pirolli, P., Chen, K., and Pitkow, J. (2001). Using information scent to model user information needs and actions and the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 490–497. ACM.
- [43] Chi, E. H., Pirolli, P., and Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 161–168. ACM.
- [44] Chowdhury, S., Gibb, F., and Landoni, M. (2011). Uncertainty in information seeking and retrieval: A study in an academic environment. *Information Processing & Management*, 47(2):157–175.
- [45] Cohen, D. W. and Cohen, D. W. (1989). *An Introduction to Hilbert Space and Quantum Logic*. Springer.
- [46] Cohen, N., Sharir, O., and Shashua, A. (2016). On the expressive power of deep learning: A tensor analysis. In *Conference on learning theory*, pages 698–728. PMLR.
- [47] Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167.
- [48] Cummins, R. and O’Riordan, C. (2006). Evolving local and global weighting schemes in information retrieval. *Information Retrieval*, 9(3):311–330.
- [49] Das, A., Kottur, S., Moura, J. M., Lee, S., and Batra, D. (2017). Learning cooperative visual dialog agents with deep reinforcement learning. In *Proceedings of the IEEE international conference on computer vision*, pages 2951–2960.
- [50] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- [51] Diaz, F., Mitra, B., and Craswell, N. (2016). Query expansion with locally-trained word embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 367–377.
- [52] Dirac, P. A. M. (1981). *The principles of quantum mechanics*. Number 27. Oxford university press.
- [53] Dorner, D. G., Gorman, G. E., and Calvert, P. J. (2014). *Information needs analysis: Principles and practice in information organizations*. Facet Publishing.
- [54] Dou, Z., Song, R., and Wen, J.-R. (2007). A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, pages 581–590.
- [55] Du, J. T. and Spink, A. (2011). Toward a web search model: Integrating multitasking, cognitive coordination, and cognitive shifts. *Journal of the American Society for Information Science and Technology*, 62(8):1446–1472.

- [56] Dumais, S. (2010). Keynote: the web changes everything: understanding and supporting people in dynamic information environments. In *International Conference on Theory and Practice of Digital Libraries*, pages 1–1. Springer.
- [57] Efron, M. and Winget, M. (2010). Query polyrepresentation for ranking retrieval systems without relevance judgments. *Journal of the American Society for Information Science and Technology*, 61(6):1081–1091.
- [58] Eliassen, S., Jørgensen, C., Mangel, M., and Giske, J. (2007). Exploration or exploitation: life expectancy changes the value of learning in foraging strategies. *Oikos*, 116(3):513–523.
- [59] Ethayarajh, K., Duvenaud, D., and Hirst, G. (2019). Understanding undesirable word embedding associations. *arXiv preprint arXiv:1908.06361*.
- [60] Fakhari, P., Rajagopal, K., Balakrishnan, S., and Busemeyer, J. (2013). Quantum inspired reinforcement learning in changing environment. *New Mathematics and Natural Computation*, 9(03):273–294.
- [61] Feher, G., Spitz, A., and Gertz, M. (2019). Retrieving multi-entity associations: An evaluation of combination modes for word embeddings. *arXiv preprint arXiv:1905.09052*.
- [62] Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Mikolov, T., et al. (2013). Devise: A deep visual-semantic embedding model. In *Advances in neural information processing systems*, pages 2121–2129.
- [63] Frommholz, I., Larsen, B., Piwowarski, B., Lalmas, M., Ingwersen, P., and Van Rijsbergen, K. (2010). Supporting polyrepresentation in a quantum-inspired geometrical retrieval framework. In *Proceedings of the third symposium on Information interaction in context*, pages 115–124. ACM.
- [64] Frommholz, I., Piwowarski, B., Lalmas, M., and Van Rijsbergen, K. (2011). Processing queries in session in a quantum-inspired ir framework. In *European Conference on Information Retrieval*, pages 751–754. Springer.
- [65] Fu, W.-T. (2008). The microstructures of social tagging: a rational model. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 229–238. ACM.
- [66] Fuhr, N. (1992). Probabilistic models in information retrieval. *The computer journal*, 35(3):243–255.
- [67] Fuhr, N. (2008). A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11(3):251–265.
- [68] Gaikwad, M. and Hoerber, O. (2019). An interactive image retrieval approach to searching for images on social media. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 173–181. ACM.
- [69] Ganguly, D., Roy, D., Mitra, M., and Jones, G. J. (2015). Word embedding based generalized language model for information retrieval. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 795–798.

- [70] Gao, D., Jin, L., Chen, B., Qiu, M., Li, P., Wei, Y., Hu, Y., and Wang, H. (2020). Fashionbert: Text and image matching with adaptive loss for cross-modal retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2251–2260.
- [71] Gardenfors, P. (2004). Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, 2(2):9–27.
- [72] Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. Hachette UK.
- [73] Goodrum, A. and Spink, A. (1999). Visual information seeking: A study of image queries on the world wide web. In *Proceedings of the ASIST Annual Meeting*, volume 36, pages 665–74.
- [74] Goodwin, J. C., Cohen, T., and Rindflesch, T. (2012). Discovery by scent: Discovery browsing system based on the information foraging theory. In *2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, pages 232–239. IEEE.
- [75] Grover, L. K. (1997). Quantum mechanics helps in searching for a needle in a haystack. *Physical review letters*, 79(2):325.
- [76] Guo, X., Wu, H., Cheng, Y., Rennie, S., Tesauro, G., and Feris, R. (2018). Dialog-based interactive image retrieval. In *Advances in Neural Information Processing Systems*, pages 678–688.
- [77] Haj-Yahia, Z., Sieg, A., and Deleris, L. A. (2019). Towards unsupervised text classification leveraging experts and word embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 371–379.
- [78] Han, X., Wu, Z., Huang, P. X., Zhang, X., Zhu, M., Li, Y., Zhao, Y., and Davis, L. S. (2017). Automatic spatially-aware fashion concept discovery. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1463–1471.
- [79] Hauff, C., Murdock, V., and Baeza-Yates, R. (2008). Improved query difficulty prediction for the web. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 439–448. ACM.
- [80] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- [81] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [82] He, R. and McAuley, J. (2016). Vbpr: visual bayesian personalized ranking from implicit feedback. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [83] Hernández-Rubio, M., Cantador, I., and Bellogín, A. (2019). A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. *User Modeling and User-Adapted Interaction*, 29(2):381–441.

- [84] Hills, T. T., Jones, M. N., and Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological review*, 119(2):431.
- [85] Hofmann, K., Li, L., Radlinski, F., et al. (2016). Online evaluation for information retrieval. *Foundations and Trends® in Information Retrieval*, 10(1):1–117.
- [86] Hosseinzadeh, M. and Wang, Y. (2020). Composed query image retrieval using locally bounded features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3596–3605.
- [87] Hu, R., Dollár, P., He, K., Darrell, T., and Girshick, R. (2018). Learning to segment every thing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4233–4241.
- [88] Hu, R., Li, S., and Liang, S. (2019). Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908.
- [89] Hu, R., Xu, H., Rohrbach, M., Feng, J., Saenko, K., and Darrell, T. (2016). Natural language object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4555–4564.
- [90] Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of documentation*, 52(1):3–50.
- [91] Ingwersen, P. and Järvelin, K. (2006). *The turn: Integration of information seeking and retrieval in context*, volume 18. Springer Science & Business Media.
- [92] Isola, P., Lim, J. J., and Adelson, E. H. (2015). Discovering states and transformations in image collections. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1383–1391.
- [93] Jaech, A. and Ostendorf, M. (2018). Personalized language model for query auto-completion. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 700–705.
- [94] Jaiswal, A. K. (2018). Investigating interactive information retrieval via information foraging theory.
- [95] Jaiswal, A. K., Holdack, G., Frommholz, I., and Liu, H. (2018). Quantum-like generalization of complex word embedding: A lightweight approach for textual classification. In Gemulla, R., Ponzetto, S. P., Bizer, C., Keuper, M., and Stuckenschmidt, H., editors, *Proceedings of the Conference "Lernen, Wissen, Daten, Analysen", LWDA 2018, Mannheim, Germany, August 22-24, 2018*, volume 2191 of *CEUR Workshop Proceedings*, pages 159–168. CEUR-WS.org.
- [96] Jaiswal, A. K., Liu, H., and Frommholz, I. (2019a). Effects of foraging in personalized content-based image recommendation. *arXiv preprint arXiv:1907.00483*.
- [97] Jaiswal, A. K., Liu, H., and Frommholz, I. (2019b). Information foraging for enhancing implicit feedback in content-based image recommendation. In Majumder, P., Mitra, M., Gangopadhyay, S., and Mehta, P., editors, *FIRE '19: Forum for Information Retrieval Evaluation, Kolkata, India, December, 2019*, pages 65–69. ACM.

- [98] Jaiswal, A. K., Liu, H., and Frommholz, I. (2020a). Reinforcement learning-driven information seeking: A quantum probabilistic approach. In Frommholz, I., Liu, H., and Melucci, M., editors, *Proceedings of the First Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2020) co-located with 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020), Xi'an, China (online only), July 30th, - 2020*, volume 2741 of *CEUR Workshop Proceedings*, pages 16–29. CEUR-WS.org.
- [99] Jaiswal, A. K., Liu, H., and Frommholz, I. (2020b). Utilising information foraging theory for user interaction with image query auto-completion. In *European Conference on Information Retrieval*, pages 666–680. Springer.
- [100] Jaiswal, A. K., Liu, H., and Frommholz, I. (2021). Semantic hilbert space for interactive image retrieval. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 307–315.
- [101] Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2007). Tag recommendations in folksonomies. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 506–514. Springer.
- [102] Ji, S., Li, G., Li, C., and Feng, J. (2009). Efficient interactive fuzzy keyword search. In *Proceedings of the 18th international conference on World wide web*, pages 371–380. ACM.
- [103] Jiang, J.-Y., Ke, Y.-Y., Chien, P.-Y., and Cheng, P.-J. (2014a). Learning user reformulation behavior for query auto-completion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 445–454. ACM.
- [104] Jiang, M., Cui, P., Wang, F., Zhu, W., and Yang, S. (2014b). Scalable recommendation with social contextual information. *IEEE Transactions on Knowledge and Data Engineering*, 26(11):2789–2802.
- [105] Jin, X., Sloan, M., and Wang, J. (2013). Interactive exploratory search for multi page search results. In *Proceedings of the 22nd international conference on World Wide Web*, pages 655–666.
- [106] Jing, Y., Zhang, X., Wu, L., Wang, J., Feng, Z., and Wang, D. (2014). Recommendation on flickr by combining community user ratings and item importance. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE.
- [107] Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. (2017). Accurately interpreting clickthrough data as implicit feedback. In *ACM SIGIR Forum*, volume 51, pages 4–11. Acm.
- [108] Johansson, R. and Pina, L. N. (2015). Embedding a semantic network in a word space. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1428–1433.
- [109] Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

- [110] Kadin, A. M. (2005). Quantum mechanics without complex numbers: A simple model for the electron wavefunction including spin. *arXiv preprint quant-ph/0502139*.
- [111] Kazemzadeh, S., Ordonez, V., Matten, M., and Berg, T. (2014). Referitgame: Referring to objects in photographs of natural scenes. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 787–798.
- [112] Kharitonov, E., Macdonald, C., Serdyukov, P., and Ounis, I. (2013). User model-based metrics for offline query suggestion evaluation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 633–642. ACM.
- [113] Khrennikov, A. and Basieva, I. (2014). Possibility to agree on disagree from quantum information and decision making. *Journal of Mathematical Psychology*, 62:1–15.
- [114] Kim, W., Hantula, D. A., and Di Benedetto, C. A. (2016). The role of foraging theory in information overload paradigm: Consumer perception of online information structures among goods and services. In *2016 Global Marketing Conference at Hong Kong*, pages 327–328.
- [115] Kiros, R., Salakhutdinov, R., and Zemel, R. S. (2014). Unifying visual-semantic embeddings with multimodal neural language models. *arXiv preprint arXiv:1411.2539*.
- [116] Kovashka, A., Parikh, D., and Grauman, K. (2015). Whittlesearch: Interactive image search with relative attribute feedback. *International Journal of Computer Vision*, 115(2):185–210.
- [117] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D. A., et al. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1):32–73.
- [118] Kwong, C. P. (2009). The mystery of square root of minus one in quantum mechanics, and its demystification. *arXiv preprint arXiv:0912.3996*.
- [119] Lefortier, D., Serdyukov, P., and De Rijke, M. (2014). Online exploration for detecting shifts in fresh intent. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 589–598. ACM.
- [120] Lespagnol, C., Mothe, J., and Ullah, M. Z. (2019). Information nutritional label and word embedding to estimate information check-worthiness. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 941–944. ACM.
- [121] Li, J., Zhang, M., Ma, W., Liu, Y., and Ma, S. (2020). A multi-level interactive lifelog search engine with user feedback. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 29–35.
- [122] Li, J., Zhang, P., Song, D., and Hou, Y. (2016). An adaptive contextual quantum language model. *Physica A: Statistical Mechanics and its Applications*, 456:51–67.

- [123] Li, Q., Li, J., Zhang, P., and Song, D. (2015). Modeling multi-query retrieval tasks using density matrix transformation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 871–874.
- [124] Li, Q., Uprety, S., Wang, B., and Song, D. (2018). Quantum-inspired complex word embedding. In *Proceedings of The Third Workshop on Representation Learning for NLP*, pages 50–57.
- [125] Li, Q., Wang, B., and Melucci, M. (2019). Cnm: An interpretable complex-valued network for matching. *arXiv preprint arXiv:1904.05298*.
- [126] Li, Y., Dong, A., Wang, H., Deng, H., Chang, Y., and Zhai, C. (2014a). A two-dimensional click model for query auto-completion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 455–464. ACM.
- [127] Li, Y., Luo, J., and Mei, T. (2014b). Personalized image recommendation for web search engine users. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE.
- [128] Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073.
- [129] Liu, H., Mulholland, P., Song, D., Uren, V., and R uger, S. (2010). Applying information foraging theory to understand user interaction with content-based image retrieval. In *Proceedings of the third symposium on Information interaction in context*, pages 135–144. ACM.
- [130] Liu, H., Mulholland, P., Song, D., Uren, V., and R uger, S. (2011). An information foraging theory based user study of an adaptive user interaction framework for content-based image retrieval. In *International Conference on Multimedia Modeling*, pages 241–251. Springer.
- [131] Loumakis, F., Stumpf, S., and Grayson, D. (2011). This image smells good: effects of image information scent in search engine results pages. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 475–484. ACM.
- [132] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6382–6393.
- [133] Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development*, 1(4):309–317.
- [134] Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
- [135] MacAvaney, S., Yates, A., Cohan, A., and Goharian, N. (2019). Cedr: Contextualized embeddings for document ranking. *arXiv preprint arXiv:1904.07094*.

- [136] Mantovani, G. (2001). The psychological construction of the internet: From information foraging to social gathering to cultural mediation. *CyberPsychology & Behavior*, 4(1):47–56.
- [137] Maxwell, D. and Azzopardi, L. (2018). Information scent, searching and stopping. In *European Conference on Information Retrieval*, pages 210–222. Springer.
- [138] McCann, B., Bradbury, J., Xiong, C., and Socher, R. (2017). Learned in translation: Contextualized word vectors. In *Advances in Neural Information Processing Systems*, pages 6294–6305.
- [139] Melucci, M. (2005). Context modeling and discovery using vector space bases. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 808–815. ACM.
- [140] Melucci, M. (2008). A basis for information retrieval in context. *ACM Transactions on Information Systems (TOIS)*, 26(3):14.
- [141] Messina, P., Dominguez, V., Parra, D., Trattner, C., and Soto, A. (2019). Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features. *User Modeling and User-Adapted Interaction*, 29(2):251–290.
- [142] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- [143] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- [144] Mitra, B. (2015). Exploring session context using distributed representations of queries and reformulations. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 3–12. ACM.
- [145] Mitra, B. and Craswell, N. (2015). Query auto-completion for rare prefixes. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1755–1758. ACM.
- [146] Mitra, B., Rosset, C., Hawking, D., Craswell, N., Diaz, F., and Yilmaz, E. (2019). Incorporating query term independence assumption for efficient retrieval and ranking using deep neural networks. *arXiv preprint arXiv:1907.03693*.
- [147] Moshfeghi, Y. and Jose, J. M. (2013). On cognition, emotion, and interaction aspects of search tasks with different search intentions. In *Proceedings of the 22nd international conference on World Wide Web*, pages 931–942. ACM.
- [148] Mu, J. and Viswanath, P. (2018). All-but-the-top: Simple and effective postprocessing for word representations. In *International Conference on Learning Representations*.
- [149] Neisser, U. (1976). Cognition and reality san francisco. *Freeman*. Newell, A.(1973). *You can't play*, 20:283–308.

- [150] Neyshabur, B., Bhojanapalli, S., McAllester, D., and Srebro, N. (2017). Exploring generalization in deep learning. In *Advances in neural information processing systems*, pages 5947–5956.
- [151] Nielsen, M. A. and Chuang, I. L. (2010). Quantum computation and quantum information.
- [152] Nogueira, R., Bulian, J., and Ciaramita, M. (2018). Learning to coordinate multiple reinforcement learning agents for diverse query reformulation. *arXiv preprint arXiv:1809.10658*.
- [153] Oard, D. W. and Kim, J. (2001). Modeling information content using observable behavior.
- [154] O’Day, V. L. and Jeffries, R. (1993). Orienteering in an information landscape: how information seekers get from here to there. In *Proceedings of the INTERACT’93 and CHI’93 conference on Human factors in computing systems*, pages 438–445.
- [155] O’Hare, N., De Juan, P., Schifanella, R., He, Y., Yin, D., and Chang, Y. (2016). Leveraging user interaction signals for web image search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 559–568. ACM.
- [156] Ong, K. (2017). Using information foraging theory to understand search behavior in different environments. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 411–413.
- [157] Ong, K., Järvelin, K., Sanderson, M., and Scholer, F. (2017). Using information scent to understand mobile and desktop web search behavior. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 295–304.
- [158] Pagán, A. and Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive science*, 43(1):e12705.
- [159] Park, D. H. and Chiba, R. (2017). A neural language model for query auto-completion. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1189–1192. ACM.
- [160] Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- [161] Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A. (2017). Film: Visual reasoning with a general conditioning layer. *arXiv preprint arXiv:1709.07871*.
- [162] Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive science*, 29(3):343–373.
- [163] Pirolli, P. (2006). The use of proximal information scent to forage for distal content on the world wide web. *Adaptive perspectives on human-technology interaction: Methods and models for cognitive engineering and human-computer interaction*, pages 247–266.

- [164] Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.
- [165] Pirolli, P. and Card, S. (1995). Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–58.
- [166] Pirolli, P. and Card, S. (1999). Information foraging. *Psychological review*, 106(4):643.
- [167] Pirolli, P., Card, S. K., and Van Der Wege, M. M. (2001). Visual information foraging in a focus+ context visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 506–513. ACM.
- [168] Pirolli, P., Card, S. K., and Van Der Wege, M. M. (2003). The effects of information scent on visual search in the hyperbolic tree browser. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(1):20–53.
- [169] Pirolli, P., Fu, W.-T., Reeder, R., and Card, S. K. (2002). A user-tracing architecture for modeling interaction with the world wide web. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 75–83. ACM.
- [170] Piwowarski, B., Frommholz, I., Lalmas, M., and Van Rijsbergen, K. (2010). What can quantum theory bring to information retrieval. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 59–68.
- [171] Piwowarski, B. and Lalmas, M. (2009). A quantum-based model for interactive information retrieval (extended version). *arXiv preprint arXiv:0906.4026*.
- [172] Plummer, P., Perea, M., and Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1):275.
- [173] Raunak, V., Gupta, V., and Metze, F. (2019). Effective dimensionality reduction for word embeddings. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pages 235–243.
- [174] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.
- [175] Rendle, S., Freudenthaler, C., Gantner, Z., and Schmidt-Thieme, L. (2009). Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press.
- [176] Rényi, A. (1955). On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungarica*, 6(3-4):285–335.
- [177] Robertson, S. E. (1977). The probability ranking principle in ir. *Journal of documentation*, 33(4):294–304.
- [178] Robertson, S. E. and Walker, S. (1999). Okapi/keenbow at trec-8. In *TREC*, volume 8, pages 151–162. Citeseer.

- [179] Rocchio, J. J. (1971). Relevance feedback in information retrieval. *The SMART retrieval system: experiments in automatic document processing*, pages 313–323.
- [180] Rose, D. E. and Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, pages 13–19.
- [181] Saleh, B. and Elgammal, A. (2015). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*.
- [182] Salle, A., Idiart, M., and Villavicencio, A. (2016a). Enhancing the lexvec distributed word representation model using positional contexts and external memory. *arXiv preprint arXiv:1606.01283*.
- [183] Salle, A., Villavicencio, A., and Idiart, M. (2016b). Matrix factorization using window sampling and negative sampling for improved word representations. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 419–424.
- [184] Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.
- [185] Sang, J. and Xu, C. (2012). Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 19–28. ACM.
- [186] Santoro, A., Raposo, D., Barrett, D. G., Malinowski, M., Pascanu, R., Battaglia, P., and Lillicrap, T. (2017). A simple neural network module for relational reasoning. In *Advances in neural information processing systems*, pages 4967–4976.
- [187] Savolainen, R. (2012). Conceptualizing information need in context.
- [188] Schnabel, T., Bennett, P. N., Dumais, S. T., and Joachims, T. (2016). Using shortlists to support decision making and improve recommender system performance. In *Proceedings of the 25th International Conference on World Wide Web*, pages 987–997. International World Wide Web Conferences Steering Committee.
- [189] Schnabel, T., Bennett, P. N., and Joachims, T. (2019). Shaping feedback data in recommender systems with interventions based on information foraging theory. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 546–554. ACM.
- [190] Schneiderman, S. (1999). Information visualization: Using vision to think.
- [191] Schütze, H., Manning, C. D., and Raghavan, P. (2008). *Introduction to information retrieval*, volume 39. Cambridge University Press Cambridge.
- [192] Seo, Y.-W. and Zhang, B.-T. (2000). A reinforcement learning agent for personalized information filtering. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 248–251. ACM.
- [193] Shao, T., Chen, H., and Chen, W. (2018). Query auto-completion based on word2vec semantic similarity. In *Journal of Physics: Conference Series*, volume 1004, page 012018. IOP Publishing.

- [194] Shen, X., Tan, B., and Zhai, C. (2005). Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824–831. ACM.
- [195] Shi, H., Li, H., Meng, F., and Wu, Q. (2018). Key-word-aware network for referring expression image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 38–54.
- [196] Shokouhi, M. (2013). Learning to personalize query auto-completion. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 103–112. ACM.
- [197] Sloan, M. and Wang, J. (2015). Dynamic information retrieval: Theoretical framework and application. In *Proceedings of the 2015 International Conference on the theory of Information Retrieval*, pages 61–70.
- [198] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.
- [199] Song, D., Lalmas, M., van Rijsbergen, K., Frommholz, I., Piwowarski, B., Wang, J., Zhang, P., Zuccon, G., Bruza, P., Arafat, S., et al. (2010). How quantum theory is developing the field of information retrieval. In *2010 AAAI Fall Symposium Series*.
- [200] Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Grue Simonsen, J., and Nie, J.-Y. (2015). A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 553–562.
- [201] Sordoni, A., Nie, J.-Y., and Bengio, Y. (2013). Modeling term dependencies with quantum language models for ir. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 653–662. ACM.
- [202] Sundar, S. S., Knobloch-Westerwick, S., and Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, 58(3):366–378.
- [203] Sutskever, I., Martens, J., and Hinton, G. E. (2011). Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1017–1024.
- [204] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- [205] Tang, Z. and Yang, G. H. (2019). Dynamic search—optimizing the game of information seeking. *arXiv preprint arXiv:1909.12425*.
- [206] Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J. (2018). Deep complex networks. In *International Conference on Learning Representations*.

- [207] Tran, V. T. and Fuhr, N. (2012). Using eye-tracking with dynamic areas of interest for analyzing interactive information retrieval. In *SIGIR*, pages 1165–1166.
- [208] Turian, J., Ratinov, L., and Bengio, Y. (2010). Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394.
- [209] Uprety, S., Dehdashti, S., Fell, L., Bruza, P., and Song, D. (2019). Modelling dynamic interactions between relevance dimensions. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 35–42. ACM.
- [210] van Rijsbergen, C. J. (1989). Towards an information logic. In *Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 77–86.
- [211] Van Rijsbergen, C. J. (2004). *The geometry of information retrieval*. Cambridge University Press.
- [212] Vig, J., Sen, S., and Riedl, J. (2009). Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 47–56. ACM.
- [213] Vijayakumar, A. K., Cogswell, M., Selvaraju, R. R., Sun, Q., Lee, S., Crandall, D., and Batra, D. (2016). Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv preprint arXiv:1610.02424*.
- [214] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164.
- [215] Vo, N., Jiang, L., Sun, C., Murphy, K., Li, L.-J., Fei-Fei, L., and Hays, J. (2019). Composing text and image for image retrieval-an empirical odyssey. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6439–6448.
- [216] Wang, B., Li, Q., Melucci, M., and Song, D. (2019). Semantic hilbert space for text representation learning. In *The World Wide Web Conference*, pages 3293–3299. ACM.
- [217] Wang, B., Yang, Y., Xu, X., Hanjalic, A., and Shen, H. T. (2017a). Adversarial cross-modal retrieval. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 154–162.
- [218] Wang, B., Zhang, P., Li, J., Song, D., Hou, Y., and Shang, Z. (2016a). Exploration of quantum interference in document relevance judgement discrepancy. *Entropy*, 18(4):144.
- [219] Wang, B., Zhao, D., Lioma, C., Li, Q., Zhang, P., and Simonsen, J. G. (2020). Encoding word order in complex embeddings. In *International Conference on Learning Representations*.
- [220] Wang, J., Song, D., and Kaliciak, L. (2010a). Tensor product of correlated textual and visual features: A quantum theory inspired image retrieval framework. In *2010 AAAI Fall Symposium Series*.

- [221] Wang, K., Gloy, N., and Li, X. (2010b). Inferring search behaviors using partially observable markov (pom) model. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 211–220. ACM.
- [222] Wang, L., Li, Y., and Lazebnik, S. (2016b). Learning deep structure-preserving image-text embeddings. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5005–5013.
- [223] Wang, P., Hou, Y., Li, J., Zhang, Y., Song, D., and Li, W. (2017b). A quasi-current representation for information needs inspired by two-state vector formalism. *Physica A: Statistical Mechanics and its Applications*, 482:627–637.
- [224] Wang, Y., Yin, D., Jie, L., Wang, P., Yamada, M., Chang, Y., and Mei, Q. (2016c). Beyond ranking: Optimizing whole-page presentation. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 103–112. ACM.
- [225] Wells, V. K. (2012). Foraging: An ecology model of consumer behaviour? *Marketing Theory*, 12(2):117–136.
- [226] Westman, S. (2009). Image users’ needs and searching behaviour. *Information Retrieval: Searching in the 21st Century*, page 63.
- [227] White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12. ACM.
- [228] White, R. W. (2014). Belief dynamics in web search. *Journal of the Association for Information Science and Technology*, 65(11):2165–2178.
- [229] White, R. W., Bennett, P. N., and Dumais, S. T. (2010). Predicting short-term interests using activity-based search context. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1009–1018. ACM.
- [230] White, R. W. and Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. *Information Processing & Management*, 43(3):685–704.
- [231] White, R. W. and Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98.
- [232] Widdows, D. and Widdows, D. (2004). *Geometry and meaning*, volume 773. CSLI publications Stanford.
- [233] Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.
- [234] Wilson, M. L. et al. (2008). Improving exploratory search interfaces: Adding value or information overload?
- [235] Wilson, T. D. (1981). On user studies and information needs. *Journal of documentation*.

- [236] Wittek, P., Lim, I. S., and Rubio-Campillo, X. (2013). Quantum probabilistic description of dealing with risk and ambiguity in foraging decisions. In *International Symposium on Quantum Interaction*, pages 296–307. Springer.
- [237] Wittek, P., Liu, Y.-H., Darányi, S., Gedeon, T., and Lim, I. S. (2016). Risk and ambiguity in information seeking: Eye gaze patterns reveal contextual behavior in dealing with uncertainty. *Frontiers in psychology*, 7:1790.
- [238] Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.
- [239] Wu, C.-C., Mei, T., Hsu, W. H., and Rui, Y. (2014a). Learning to personalize trending image search suggestion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 727–736. ACM.
- [240] Wu, W.-C., Kelly, D., and Sud, A. (2014b). Using information scent and need for cognition to understand online search behavior. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 557–566. ACM.
- [241] Xie, M., Hou, Y., Zhang, P., Li, J., Li, W., and Song, D. (2015). Modeling quantum entanglements in quantum language models. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- [242] Xie, X., Mao, J., Liu, Y., de Rijke, M., Ai, Q., Huang, Y., Zhang, M., and Ma, S. (2019). Improving web image search with contextual information. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1683–1692.
- [243] Xu, D., Liu, Y., Zhang, M., Ma, S., and Ru, L. (2012). Incorporating revisiting behaviors into click models. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 303–312. ACM.
- [244] Yue, Y. and Joachims, T. (2009). Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1201–1208. ACM.
- [245] Zendel, O., Shtok, A., Raiber, F., Kurland, O., and Culpepper, J. S. (2019). Information needs, queries, and query performance prediction. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 395–404.
- [246] Zhai, C. and Lafferty, J. (2002). Two-stage language models for information retrieval. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 49–56. ACM.
- [247] Zhang, B.-T. and Seo, Y.-W. (2001). Personalized web-document filtering using reinforcement learning. *Applied Artificial Intelligence*, 15(7):665–685.
- [248] Zhang, L., Qi, G.-J., Wang, L., and Luo, J. (2019a). Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2547–2555.

- [249] Zhang, L., Zhang, P., Ma, X., Gu, S., Su, Z., and Song, D. (2019b). A generalized language model in tensor space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7450–7458.
- [250] Zhang, P., Niu, J., Su, Z., Wang, B., Ma, L., and Song, D. (2018a). End-to-end quantum-like language models with application to question answering. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [251] Zhang, P., Su, Z., Zhang, L., Wang, B., and Song, D. (2018b). A quantum many-body wave function inspired language modeling approach. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1303–1312.
- [252] Zhang, Y., Li, Q., Song, D., Zhang, P., and Wang, P. (2019c). Quantum-inspired interactive networks for conversational sentiment analysis.
- [253] Zhang, Y., Song, D., Zhang, P., Wang, P., Li, J., Li, X., and Wang, B. (2018c). A quantum-inspired multimodal sentiment analysis framework. *Theoretical Computer Science*, 752:21–40.
- [254] Zhen, L., Hu, P., Wang, X., and Peng, D. (2019). Deep supervised cross-modal retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10394–10403.
- [255] Zhou, J. and Agichtein, E. (2020). Rlirank: Learning to rank with reinforcement learning for dynamic search. In *Proceedings of The Web Conference 2020*, pages 2842–2848.
- [256] Zuccon, G. (2012). *Document ranking with quantum probabilities*. PhD thesis, University of Glasgow.

Appendix A

Ethical Approval

We describe the details of ethical approval procedure for the pilot user study reported in this dissertation, applications to the University of Bedfordshire Research Ethics Board and participant consent forms.

A.1 Application Form for Pilot User Study

Here, I include the actual application form that details the ethical approval process.

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Postgraduate Research Students)

When completing this form please ensure that you read and comply with the following:

Researchers must demonstrate clear understanding of an engagement with the following:

1. *Integrity* - The research has been carried out in a rigorous and professional manner and due credit has been attributed to all parties involved.
2. *Plagiarism* - Proper acknowledgement has been given to the authorship of data and ideas.
3. *Conflicts of Interest* - All financial and professional conflicts of interest have been properly identified and declared.
4. *Data Handling* - The research draws upon effective record keeping, proper storage of data in line with confidentiality, statute and University policy.
5. *Ethical Procedures* - Proper consideration has been given to all ethical issues and appropriate approval sought and received from all relevant stakeholders. In addition the research should conform to professional codes of conduct where appropriate.
6. *Supervision* - Effective management and supervision of staff and student for whom the researcher(s) is/are responsible
7. *Health and Safety*- Proper training on health and safety issues has been received and completed by all involved parties. Health and safety issues have been identified and appropriate assessment and action have been undertaken.

The **Research Institutes** are responsible for ensuring that all researchers abide by the above. It is anticipated that ethical approval will be granted by each Research Institute. Each Research Institute will give guidance and approval on ethical procedures and ensure they conform to the requirements of relevant professional bodies. As such Research Institutes are required to provide the University Research Ethics Committee with details of their procedures for ensuring adherence to relevant ethical requirements. This applies to any research whether it be, or not, likely to raise ethical issues. Research proposals involving vulnerable groups; sensitive topics; groups requiring gatekeeper permission; deception or without full informed consent; use of personal/confidential information; subjects in stress, anxiety, humiliation or intrusive interventions must be referred to the University Research Ethics Committee.

Research projects involving participants in the NHS will be submitted through the NHS National Research Ethics Service (NRES). The University Research Ethics Committee will normally accept the judgement of NRES (it will never approve a proposal that has been rejected by NRES), however NRES approval will need to be verified before research can commence and the nature of the research will need to be verified.

Where work is conducted in collaboration with other institutions ethical approval by the University and the collaborating partner(s) will be required.

The **University Research Ethics Committee** is a sub-committee of the Academic Board and is chaired by a member of the Vice Chancellor's Executive Group, appointed by the Vice-Chancellor and includes members external to the University

Research Misconduct: Allegations of Research Misconduct against staff or post graduate (non-taught) research students should be made to the Director of Research Development.

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Annex to RS1 form)

SECTION A To be completed by the candidate

Registration No: **1808389**

Candidate: **Amit Kumar Jaiswal**

Degree of: **Ph.D.**

Research Institute: **Institute for Research in Applicable Computing**

Research Topic: **A Quantum Model for Interactive Information Retrieval based on Information Foraging Theory**

External Funding: **QUARTZ Project funded by European Union's Horizon 2020 under the Marie Sklodowska-Curie grant agreement No. 721321**

The candidate is required to summarise in the box below the ethical issues involved in the research proposal and how they will be addressed. In any proposal involving human participants the following should be provided:

- clear explanation of how informed consent will be obtained,
- how will confidentiality and anonymity be observed,
- how will the nature of the research, its purpose and the means of dissemination of the outcomes be communicated to participants,
- how personal data will be stored and secured
- if participants are being placed under any form of stress (physical or mental) identify what steps are being taken to minimise risk

If protocols are being used that have already received University Research Ethics Committee (UREC) ethical approval then please specify. Roles of any collaborating institutions should be clearly identified. Reference should be made to the appropriate professional body code of practice.

Possible ethical issues and informed consent will be obtained with the following justifications:

- **Participants**: Participants will be recruited from the University of Bedfordshire to participate in the experiment. To ensure that all participants fit in the study requirement, we will recruit participants from the all over the institute within University of Bedfordshire (undergraduate/postgraduate students) with adequate knowledge of information retrieval (or web search).
- **User consent**: All participants will be provided a consent form before the study take place, and at any point during data collection. If participant decides to withdraw from the study at any point of time throughout the study, this will be respected.
- **User privacy**: Copies of written consent forms and all participant's contact details will be kept confidential and held in a file identified only by a anonymized number and stored in a locked cabinet in my office.
- **The collected data** will not be shared with others during the study without user's consent. It will be used ONLY for the research purpose, and it can be provided to researchers/participants on request.
- **Data handling**: Collected data will be stored on a secure protected computer within Luton campus and can be only accessed by me and my academic supervisors. As per University of Bedfordshire and GDPR requirements/guidelines, all raw and analysed data will be kept for few months after the conclusion/completion of the researcher's study.
- **Apparatus**: To ensure a safe environment is created, the experiments will be conducted and held at Luton campus in the room F203 (PhD workspace) which is safe and have no health issues.
- **Additional emphasis and clarity** will be made to participants involved in the study, on the responsibilities of the participant's information, the time involved, how data will be analysed and protected.

Answer the following question by deleting as appropriate:

1. Does the study involve vulnerable participants or those unable to give informed consent (e.g. children, people with learning disabilities, your own students)?

No

If **YES**: Have/will Researchers be DBS checked?

No

2. Will the study require permission of a gatekeeper for access to participants (e.g. schools, self-help groups, residential homes)?

No

3. Will it be necessary for participants to be involved without consent (e.g. covert observation in non-public places)?

No

4. Will the study involve sensitive topics (e.g. sexual activity, substance abuse)?

No

5. Will blood or tissue samples be taken from participants?

No

6. Will the research involve intrusive interventions (e.g. drugs, hypnosis, physical exercise)?

No

7. Will financial or other inducements be offered to participants (except reasonable expenses)?

No

8. Will the research investigate any aspect of illegal activity?

No

9. Will participants be stressed beyond what is normal for them?

No

10. Will the study involve participants from the NHS (e.g. patients) or participants who fall under the requirements of the Mental Capacity Act 2005?

No

If you have answered yes to any of the above questions or if you consider that there are other significant ethical issues then details should be included in your summary above. If you have answered yes to Question 1 then a clear justification for the importance of the research must be provided.

*Please note if the answer to Question 10 is yes then the proposal should be submitted through **NHS research ethics approval procedures** to the appropriate **NRES**. The UREC should be informed of the outcome.

Checklist of documents which should be included:


Project proposal (with details of methodology) & source of funding	√
Documentation seeking informed consent (if appropriate)	√
Information sheet for participants (if appropriate)	√
Questionnaire (if appropriate)	√

(Tick as appropriate)

Applicant declaration

I understand that I cannot collect any data until the application referred to in this form has been approved by all relevant parties. I agree to carry out the research in the manner specified and comply with the statement of ethical requirements on page 1 of this form. If I make any changes to the approved method I will seek further ethical approval for any changes.

Signature of Applicant:



Date: 30-10-2019

Signature of Director of Studies:



Date: 30 October 2019

This form together with a copy of the research proposal should be submitted to the Research Institute Director for consideration by the Research Institute Ethics Committee/Panel

Note you cannot commence collection of research data until this form has been approved

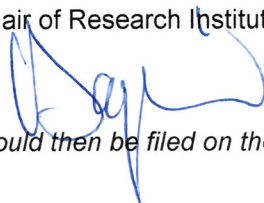
SECTION B To be completed by the Research Institute Ethics Committee:

Comments:

Approved



Signature Chair of Research Institute Ethics Committee:



Date:

31/10/2019

This form should then be filed on the student's record

If in the judgement of the committee there are significant ethical issues for which there is not agreed practice, then further ethical consideration is required before approval can be given and the proposal with the committee's comments should be forwarded to the secretary of the UREC for consideration.

There are significant ethical issues which require further guidance

Signature Chair of Research Institute Ethics Committee:

Date:

This form together with the recommendation and a copy of the research proposal should then be submitted to the University Research Ethics Committee

A.2 Participant Consent Form

Since the nature of my pilot user study were conducted on our designed content-based image recommender system, the vital thing to consider was to record the indicated consent form of the participants.

Participant Consent Form

Research Title: **Investigating Content-based Image Recommender System based on Information Foraging Theory**

Research Student: **Amit Kumar Jaiswal**

Please read and sign this form.

- You will be shortly informed about the aims of the whole study.
- You will be given 3 - 5 minutes to explore the recommender system (would be roughly 3 minutes).
- You will be given explanations to this study of food and ArtWork recommendation.
- Our content-based image recommender system follows Pinterest visual search.
- You will perform an exploratory and a recommendation task.
- The recommendation task requires ten iterations to follow through.
- The entire study relies on one dedicated interface with cascaded options to recommendation.
- The duration time of the overall experiment is about 25 minutes to be completed.
- Your personal information will be kept confidentially.
- The observation data is only for research purpose.
- Your participation in this research is voluntary and you may withdraw at any time.

This study is about investigating the user preferences via implicit feedback in Content-based Image Recommender system with the usage of Information Foraging Theory. Also, to model their preferences in order to improve selection of recommended items. The role of Information Foraging Theory is to characterise the interactive elements by means of their constructs in order for user to locate the recommended items.

Your participation in this study is voluntary. All information will remain strictly confidential. The descriptions and findings would be useful to improve the whole model via user interface functionality. However, any personally identifiable information will not be used at any point or time. You can withdraw your consent to the experiment and stop participating at any time.



We will need your email address for the purpose of contacting you for any further evaluation of the application.

I have read and understood the information on this form and had all of my questions answered.

Email address: _____

Signature: _____

Date: _____

Appendix B

Datasets

My thesis employs several publically available datasets which contain images and textual queries.

B.1 Access Link

In light of the substantial scale of the entirety of the data sets, I transferred them to a cloud-based hosting service on OneDrive in a compressed format at <https://bit.ly/thesisdataalt>, which may be accessed at the following location. In case the above link is unavailable, then the dataset can be accessible at <https://bit.ly/PhDthesisdata>.

