# Quantum-Inspired Neural Language Representation, Matching and Understanding

**Other titles in Foundations and Trends® in Information Retrieval**

*Fairness in Information Access Systems*
Michael D. Ekstrand, Anubrata Das, Robin Burke and Fernando Diaz
ISBN: 978-1-63828-040-8

*Deep Learning for Dialogue Systems: Chit-Chat and Beyond*
Rui Yan, Juntao Li and Zhou Yu
ISBN: 978-1-63828-022-4

*Search Interface Design and Evaluation*
Chang Liu, Ying-Hsang Liu, Jingjing Liu and Ralf Bierig
ISBN: 978-1-68083-922-7

*Psychology-informed Recommender Systems*
Elisabeth Lex, Dominik Kowald, Paul Seitlinger, Thi Ngoc Trang Tran, Alexander Felfernig and Markus Schedl
ISBN: 978-1-68083-844-2

*Search and Discovery in Personal Email Collections*
Michael Bendersky, Xuanhui Wang, Marc Najork and Donald Metzler
ISBN: 978-1-68083-838-1

# Quantum-Inspired Neural Language Representation, Matching and Understanding

**Peng Zhang**
Tianjin University
pzhang@tju.edu.cn

**Hui Gao**
Tianjin University
hui_gao@tju.edu.cn

**Jing Zhang**
Tianjin University
zhang_jing@tju.edu.cn

**Dawei Song**
Beijing Institute of Technology
dawei.song2010@gmail.com

now
the essence of knowledge
Boston — Delft

# Foundations and Trends® in Information Retrieval

*Published, sold and distributed by:*
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

*Outside North America:*
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

P. Zhang *et al.*. *Quantum-Inspired Neural Language Representation, Matching and Understanding*. Foundations and Trends® in Information Retrieval, vol. 16, no. 4-5, pp. 318–509, 2023.

ISBN: 978-1-63828-205-1

**Foundations and Trends® in Information Retrieval**
Volume 16, Issue 4-5, 2023

# Editorial Board

# Editorial Scope

## Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

## Information for Librarians

Foundations and Trends® in Information Retrieval, 2023, Volume 16, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

# Contents

# Quantum-Inspired Neural Language Representation, Matching and Understanding

Peng Zhang[1], Hui Gao[2], Jing Zhang[3] and Dawei Song[4]

[1] *College of Intelligence and Computing, Tianjin University, China; pzhang@tju.edu.cn*

[2] *College of Intelligence and Computing, Tianjin University, China; hui_gao@tju.edu.cn*

[3] *College of Intelligence and Computing, Tianjin University, China; zhang_jing@tju.edu.cn*

[4] *School of Computer Science and Technology, Beijing Institute of Technology, China; dawei.song2010@gmail.com*

ABSTRACT

The introduction of Quantum Theory (QT) provides a unified mathematical framework for Information Retrieval (IR). Compared with the classical IR framework, the quantum-inspired IR framework is based on user-centered modeling methods to model non-classical cognitive phenomena in human relevance judgment in the IR process. With the increase of data and computing resources, neural IR methods have been applied to the text matching and understanding task of IR. Neural networks have a strong learning ability of effective representation and generalization of matching patterns from raw data. However, these methods show some unavoidable defects, such as the inability to model user cognitive phenomena, large number of model parameters

Peng Zhang, Hui Gao, Jing Zhang and Dawei Song (2023), "Quantum-Inspired Neural Language Representation, Matching and Understanding", Foundations and Trends® in Information Retrieval: Vol. 16, No. 4-5, pp 318–509. DOI: 10.1561/1500000091.

and the "black box" characteristics of network structure. These problems greatly limit the development of neural IR and related fields. Although the quantum-inspired retrieval framework can theoretically solve the above problems, it is faced with problems such as poor model efficiency and difficulty in integrating with neural network, which lead to a huge gap between QT and neural network modeling.

This review gives a systematic introduction to quantum-inspired neural IR, including quantum-inspired neural language representation, matching and understanding. This is not only helpful to non-classical phenomena modeling in IR but also to break the theoretical bottleneck of neural networks and design more transparent neural IR models. We introduce the language representation method based on QT and the quantum-inspired text matching and decision making model under neural network, which shows its theoretical advantages in document ranking, relevance matching, multimodal IR, and can be integrated with neural networks to jointly promote the development of IR. The latest progress of quantum language understanding is introduced and further topics on QT and language modeling provide readers with more materials for thinking.

# 1

# Introduction

Retrieval is the mainstream method for humans to obtain information. The library retrieval enables us to quickly find relevant information only based on keywords, topics, *etc.*, so that we can further think and develop on the shoulders of giants. Driven by the fourth industrial revolution, the rapid development of internet technology has produced more and more complex and diverse information, bringing a new challenge to retrieval. At this time, the traditional manual retrieval technology can no longer help us to obtain knowledge efficiently from massive data sets. To this end, humans have developed *Information Retrieval (IR)* that depends on computers, which is regarded as the process of "retrieving documents related to users" (Van Rijsbergen, 1979).

Relevance is an important metric to evaluate the function and effect of IR system, and is also a central theme and important research problem in IR. At present, the relevance research of IR is mainly divided into two camps (Chen, 2020): *Computer Science* and *Library and Information Science*. The research of computer science is called *system-centered school*, which focuses on the construction of retrieval system, text representation learning, design of matching function and evaluation of retrieval performance, and calculates the relevance be-

tween query and document from a system perspective. The research of library and information science focuses on the influence of users' cognition, interaction and context on retrieval results in the process of user retrieval, which is called *user-centered school*. It evaluates the relevance of documents from the user's perspective (not just query). The two schools carry on the relevance research from different perspectives, and jointly promote the development of the IR field. However, due to the gap between user cognition and system modeling, it is difficult to build a mathematical framework that can bridge such a conceptual gap.

In the process of relevance judgment, from the system perspective, relevance is a definite concept, and the "relevance" and "irrelevance" of documents cannot exist at the same time. From the user perspective, relevance is an uncertain concept, and both "relevance" and "irrelevance" of documents can exist simultaneously at the same time. Before the user judgment, the document is in the superposition state of "relevance" and "irrelevance" with different probabilities. When the user makes a judgment , a more relevant document is selected. At this time, the relevance of a document collapses from an uncertain to a certain state. This uncertain relevance judgment process is obviously more consistent with the user's cognition. Similarly, there is semantic uncertainty in word semantic understanding. The understanding of word meanings from the user's perspective is usually based on rich prior knowledge and cognitive ability. Without considering the context, a word will be naturally associated with multiple meanings, so a word can be expressed as a superposition of different word meanings.

In computer science, neural IR models (Manning, 2016; Craswell, 2017; Kenter *et al.*, 2017; Mitra and Craswell, 2017) have shown better performance than traditional IR models. Specifically, neural IR refers to the application of shallow or deep neural networks in IR tasks (Mitra and Craswell, 2018). Many *Natural Language Processing (NLP)* tasks based on neural networks are not IR tasks (*e.g.*, neural language modeling, machine translation and named entity recognition), but they can still be used as part of IR (*e.g.*, extracting semantic information of queries and documents through language modeling tasks). Based on the data driven idea, neural IR models use the representation learning ability of neural networks to learn the required features from large-scale complex data

without hand-crafted feature, and effectively improve the performance of the IR model through the back-propagation strategy (Guo *et al.*, 2016b; Pang *et al.*, 2016; Dai *et al.*, 2018). However, although the technical advantages brought by the neural network improve the upper bound of the performance of the IR model, this method is still a system-centered modeling method, which can not fundamentally solve the problem that puzzles IR, *i.e.*, the difficulty of modeling user cognition.

*Quantum Information Retrieval (QIR)* tries to solve the problem of neural IR from the user's perspective, which uses *Quantum Theory (QT)* and its mathematical framework to model non-classical phenomena in IR (Uprety *et al.*, 2020). Different from the system-centered modeling method of neural IR, QIR follows the user-centered modeling method (Ingwersen, 1992; Ingwersen, 1996; Belkin *et al.*, 1982; Saracevic, 1996), and its purpose is to build a retrieval system that is more in line with human cognition. To achieve this goal, QIR works to explore the influence of user's cognition, emotion, environment and other factors on the relevance judgment process, and quantify these factors as non-classical phenomena that can be modeled by QT.

QIR originates from the consideration of subjective relevance factors, which makes the IR process (semantic understanding, information retrieval, *etc.*) full of "uncertainty", and these uncertain subjective factors bring non-classical phenomena (Song *et al.*, 2010). In the system-centered model, information needs are reflected through queries and will not change, so document relevance is certain; while in the user-centered model, information needs are related to subjective factors such as cognition and emotion (Miwa, 2018). It could change in the retrieval process, and the relevance of documents is uncertain. We can find that different users of the same query may also get different retrieval results, that is, the final relevance in QIR not only depends on the query, but also affected by user cognition. In recommender systems, user preference is also a manifestation of uncertainty. Preference factors such as category, brand and seller would affect each other, and group preferences would also affect individual preferences. Wang *et al.* (2019d) and Wang *et al.* (2019c) found that QT (tensor product, quantum many-body wave function, *etc.*) can effectively improve model accuracy by learning user preferences interactively. We call the phenomenon caused by user cog-

nition as non-classical phenomena (*e.g.*, order effect, text interference, cognitive decision-making, *etc.*), which are often ignored by traditional IR models and have not been considered in the relevance retrieval process.

Later, QIR starts to solve the common problems of classic retrieval such as document ranking. At this stage, the research process of QIR is further improved, which comes from casual discovery and intuitive inference, and then through the exploration of quantized theoretical explanations and mathematical modeling to solve problems that cannot be solved under the classical framework. For example, the Probability Ranking Principle (PRP) of IR model assumes that documents are independent of each other, *i.e.*, the relevance of document $A$ will not affect the relevance of document $B$. However, Zuccon *et al.* (2009b) found that when judging the relevance of document $A$ and $B$ from the user's perspective, they will be affected by the dependent information $I_{AB}$ between documents (order effect), which is a non-classical phenomenon in document ranking. Therefore, Zuccon *et al.* (2009b) modeled the interaction between documents based on quantum interference theory, derived the document dependent information $I_{AB}$, and proposed a more sensitive Quantum Probability Ranking Principle (QPRP), which is an extension of the classical PRP. As we can see, QIR has sufficient research motivation, solid theoretical foundation and feasible technical scheme. At present, the application of QT in IR involves interactive IR (Ruthven, 2010), dynamic IR (Yang *et al.*, 2015a), cognitive IR (Sutcliffe and Ennis, 1998) and so on. It models the general process of IR from the perspective of users, which is not only the promotion of neural IR, but also a breakthrough to make the field of IR in line with the quantum era.

Although QIR is expected to solve the problem of user cognitive modeling, researchers find that the experimental performance of QIR models is often poor (Zhang *et al.*, 2018a), which means that we need to make full use of the modeling ability of neural IR. The retrieval framework of "quantum + neural IR" combines the advantages of quantum mechanics and neural network: on the one hand, quantum mechanics can provide a way to model non-classical phenomena by user cognition in IR, and try to explain the semantic modeling theory

under neural network (Platonov *et al.*, 2018; Surov *et al.*, 2021; Aerts and Beltran, 2022); on the other hand, neural network architecture provides strong learning ability for the QIR model to guarantee its good experimental performance in IR tasks (McClean *et al.*, 2018; Killoran *et al.*, 2019; Beer *et al.*, 2020). For example, in the early research of quantum language model, Sordoni *et al.* (2013) extended the statistical language model to the quantum language model. Although the term dependency was effectively modeled, the MRR[1] of WIKIQA[2] is 51%. Zhang *et al.* (2018a) extended the quantum language model to the neural network architecture, and the model performance was greatly improved from 51% to 66%. The combination of neural networks and QT makes QIR more effective in practice.

As a branch of IR, the retrieval model based on statistical language model demonstrates the effectiveness of language understanding for IR tasks. Therefore, the early QIR work was based on tasks such as IR and language modeling, and representative works has shown that QT and language modeling have a more essential connection (Socher *et al.*, 2013; Zhang *et al.*, 2018b). Therefore, this monograph introduces the latest progress of QT in neural IR and language modeling. As shown in Figure 1.1, quantum-inspired neural language modeling and IR have achieved many breakthroughs in language representation, matching and understanding (Zhang *et al.*, 2018a; Zhang *et al.*, 2018b; Jiang *et al.*, 2020; Ma *et al.*, 2019). Below, we will detail the intuitions and advantages of quantum-inspired neural IR and language modeling in terms of language representation, matching, and understanding.

### Quantum Language Representation

In language representation, semantic uncertainty leads to polysemy, which can not be explicitly represented by the pre-trained word embedding. Compared with neural IR, words are represented as quantum superposition states in QIR. Just as its name implies, the superposition

[1]MRR (Mean Reciprocal Rank) is a evaluation metric for Question Answering (QA) task.

[2]Common dataset for QA task.

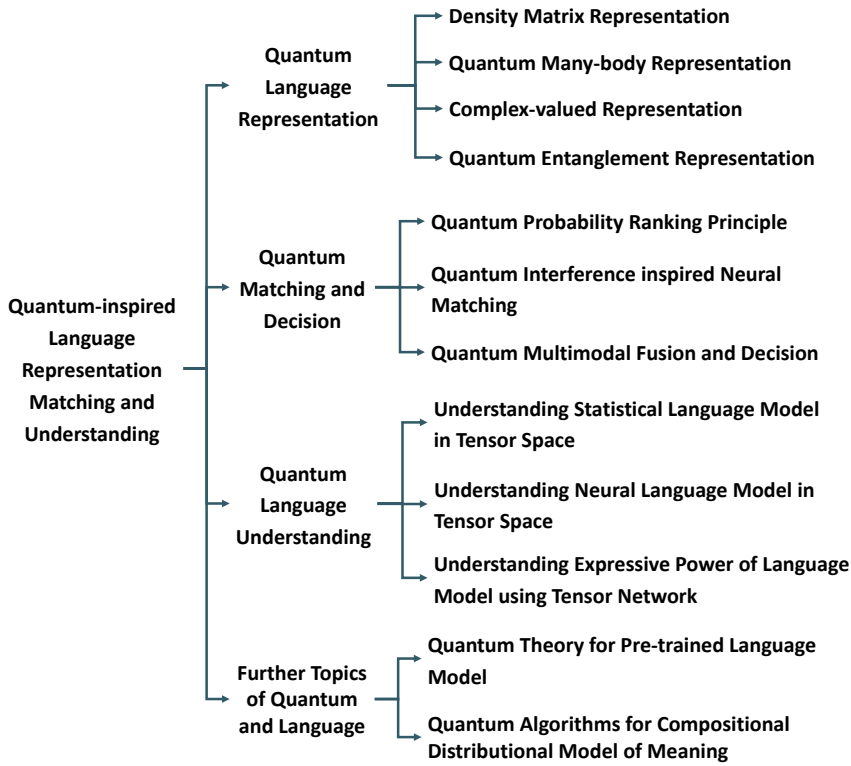



**Figure 1.1:** Structure of the quantum-inspired representation, matching and understanding for natural language and Information Retrieval.

state[3] encodes different word meanings, indicating that the word at the current moment may represent any possible meanings, and different word meanings are represented by different basis vectors. This explicit representation helps the model understand word information according to context. For example, the word "queen" can be understood as a royal, a rock band, a chess piece, or a playing card, so the word can be expressed as the superposition state of four basis vectors. When the context of the word "queen" is different, the specific meaning represented

[3]Every quantum state can be represented as a combination of two or more other distinct states (please refer to Section 2.2.1 for details).

by the current word can be determined through quantum measurement[4] (the probability of the context representation is projected to the basis vectors). This example will be detailed in Section 2.

Further, complex semantic modeling is the core problem of language representation. Traditional language modeling methods use one-hot vectors or pre-trained word embedding for word representation, and then use $n$-gram language models to model sentence semantics. However, when the value of $n$ is too large, the computational complexity increases exponentially and the neural network cannot handle it. QT is based on Hilbert space for complex semantic representation, representing words as quantum superposition states, and modeling sentence semantics in high-dimensional space through quantum many-body theory. Compared with neural networks, the tensor network used in quantum many-body theory can model high-dimensional semantic features, and has a series of tensor decomposition techniques (CP decomposition, TT decomposition, *etc.*) to facilitate the optimization calculation of high-dimensional semantics, which further improves the ability of language models to model complex relationships.

### Quantum Matching and Decision

In the process of semantic matching, the uncertainty of user cognition leads to the interaction between matching words, while the traditional IR model ignores this non-classical phenomenon and assumes the independence between matching words. The current mainstream neural IR model, such as DRMM (Pang *et al.*, 2016), formalizes the IR process as a text matching task, and its matching process is divided into two steps: the first step calculates the local relevance score between the query and the document; the second step is to obtain the final relevance score of the document by weighted summation. Compared with neural IR, QIR regards the matching process of query and document as a whole, which uses quantum interference theory to derive semantic interference term in the matching process, and embeds this component into the

[4]Quantum measurement is the testing or manipulation of a physical system to produce a numerical result, which is usually probabilistic (please refer to Section 2.2.2 for details).

neural network architecture, which is helpful to model the user cognitive information in the matching process.

In the multimodal semantic matching task, a key issue is to effectively utilize and combine multiple information from different sources from the user's perspective, so as to bridge the "semantic gap" between low-level features and high-level semantic. Compared with neural IR, QIR represents the information of different modalities in the unified Hilbert space to achieve high-dimensional semantic extraction from the perspective of user cognitive modeling, simulates the user's multimodal decision-making process through quantum interference theory, and improves the performance and interpretability of the model.

### Quantum Language Understanding

The main purpose of language modeling is to learn the joint probability function of a sequence of words. When the context of a natural language unit (such as a word) is unknown, the language model forms a parameter space that grows exponentially with sentence size (Zhang *et al.*, 2019a). Therefore, how to fully understand the parameter space of language modeling technology and interpret the internal computing mechanism of language modeling has attracted extensive attention. Taking the most classic $n$-gram language model as an example, as the number of words required to estimate the probability of a sentence increases, the parameter space of the model is difficult to represent formally. For neural language models, the parameter space is more of a black box, and its interpretability is weak, which hinders the improvement of the model's expressive ability. In addition, the expressiveness of the model is difficult to measure and analyze, resulting in the strong reliance of the model's architectural design and hyper-parameters on manual tuning.

To address the above issues, researchers have introduced QT to understand natural language modeling techniques and advance the development of language modeling interpretability. In the field of machine learning, interpretability is mainly divided into two types, one is the interpretability of the internal mechanism of the model, and the other is the interpretability of the output of the model (*i.e.*, Post-Hoc Interpretability). The interpretability based on the QT belongs

to the understanding of general language modeling techniques, that is, interpreting the internal mechanisms and parameter spaces of language models. In this monograph, we describe several ways to improve the interpretability. First, we understand the classical statistical language model in Hilbert space, formalize the parameter space of the model with high-order tensors, and derive a generalized $n$-gram language model. Next, by tensor decomposition of the probability distribution space of the statistical language model, we effectively represent the parameter space of the recurrent neural network and the convolutional neural network, interpreting and improving the expressive power of the neural language model. Finally, on the basis of the above theories, we measure the expressive power of the language model by using the separate rank and entanglement entropy, and leverage this to guide the setting of important hyper-parameters of the model to reduce the cost of manual parameter tuning. This series of studies have greatly broadened the perspective of understanding natural language, and embodies the important role of QT in the current field of machine learning interpretability.

As mentioned above, although the quantum theory is helpful for language representation, matching and understanding, most of these works are limited to the analogy between language phenomena and quantum theory. It has to be admitted that it is challenging to strictly prove the existence of quantum phenomena in human language. The quantum-like idea in this monograph is aimed at using quantum theory to understand the inherent uncertainty and cognitive mechanism in human language. We look forward to exploring the more essential relationship between quantum theory and human language, with the help of in-depth and sound research in the future.

The monograph review is organized as follows. In Section 2, we describe the commonly used quantum mathematical framework and some early QIR models. Section 3 introduces various language representation methods based on QT. Section 4 introduces the quantum matching and fusion model under the neural network architecture. Section 5 introduces language understanding based on QT and tensor network. Section 6 introduces the further topics of interdisciplinary research over quantum and language, including pre-training language model and compositional distribution model. Section 7 introduces the benchmarks involved in

this overview (including dataset, evaluation metric and performance). In Section 8, we briefly summarize the existing achievements and limitation of QIR, and look forward to the future work.

# Acknowledgements

We would like to thank Professor Yuexian Hou at Tianjin University, and all anonymous reviewers who reviewed the work and provided insightful comments.

This work is supported in part by the Natural Science Foundation of China (grant No.62276188 and No.61876129), TJU-Wenge joint laboratory funding, Tianjin Research Innovation Project for Postgraduate Students (grant No. 2021YJSB167), and MindSpore.[1]

[1] https://www.mindspore.cn/.

# Appendices

# A

# Abbreviation Index

**Table A.1:** Abbreviation index

| Full Name | Abbreviation |
|---|---|
| Artificial Intelligence | AI |
| Complex-valued Network for Matching | CNM |
| Compositional Distributional Model of Meaning | CSC |
| Convolutional Arithmetic Circuit | ConvAC |
| Density Matrix-based Convolutional Neural Network | DM-CNN |
| Hierarchical Tucker | HT |
| Information Retrieval | IR |
| Law of Total Probability | LTP |
| Machine Learning | ML |
| Maximum Likelihood Estimation | MLE |
| Matrix Product State | MPS |
| Matrix Product Operator | MPO |
| Natural Language Processing | NLP |
| Neural Network based Quantum-like Language Models | NNQLM |
| Probability Ranking Principle | PRP |
| Quantum Theory | QT |
| Quantum Interference Inspired Neural Matching Model | QINM |
| Quantum Information Retrieval | QIR |
| Quantum Language Model | QLM |
| Quantum Many-body Wave Function inspired Language Modeling | QMWF-LM |
| Quantum Language Model with Entanglement Embedding | QLM-EE |
| Quantum Probability Ranking Principle | QPRP |
| Tensor Space Language Model | TSLM |
| Text classification based on Tensor Network | TextTN |
| Tensor Network | TN |
| Tensor Train | TT |

# B

# The Fundament of Quantum Theory

## B.1 Quantum Measurement

Quantum mechanics describes a quantum system by a state vector in Hilbert space, which contains all information of the quantum system. However, the state vector is not a quantity that can be directly observed in experiments. It is necessary to introduce quantum measurement to obtain observables. In quantum measurement, each observable corresponds to a set of eigenstates and eigenvalues, and the result of observation can only be one of these eigenvalues. Specifically, by decohering the quantum system (degenerating to a classical system), the state also collapses from the superposition state to a certain eigenstate, and the observed result can be obtained.

The Born's rule (also called Born rule) provides a link between the mathematical formalism and experiment of quantum theory based on the probabilistic interpretation of the wave function. It claims that the measurement of a quantum system will produce the probability of a given result. At a given time, the particle may appear anywhere. The probability of the particle appearing at a certain point is determined by Born's rule. The wave function is the probability wave used to describe this process, and the probability of appearing at a certain position

depends on the size of the area covered by the wave function around the position (the shaded part in Figure B.1).

$A$ is an observable, represented by a self-adjoint operator on a Hilbert space, and $A$ has an orthonormal basis of eigenvectors $e_i$ with corresponding eigenvalues $\phi_i$, *i.e.*, $Ae_i = \phi_i e_i$. The Born rule states that if the quantum system is in a state value $\Phi$, then the probability $P(A = \phi_i|\Phi)$ that the eigenvalue $\phi_i$ of $A$ is found when $A$ is measured is (Landsman, 2009):

$$P\left(A = \phi_i|\Phi\right) = |\langle e_i, \Phi\rangle|^2 \tag{B.1}$$

where $\langle\ ,\ \rangle$is an operation that represents the inner product.

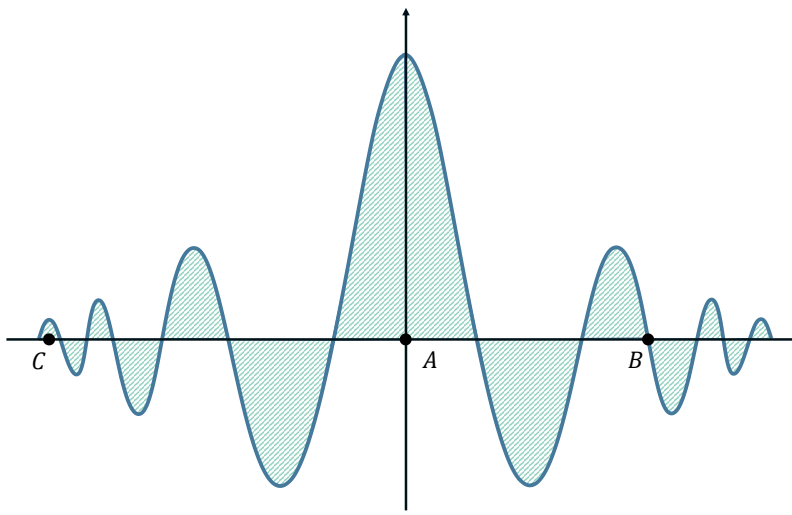



**Figure B.1:** Probability interpretation of the wave function: according to the distribution of the wave function in the figure, the probability of particles appearing around point $A$ is the highest, the probability of appearing around point $C$ is very small, and the particle cannot appear at point $B$.

Gleason's theorem is an important theory in quantum logic, which shows that in Hilbert space, the probability of obtaining a specific result of a given measurement conforms to Born's rule (Marinescu and Marinescu, 2012). In the mathematical foundations of quantum mechanics, Gleason's theorem is proposed to determine all measures

on the closed subspaces of a Hilbert space. This theory expresses all measures on the closed subspace as a function $\mu$, which assigns a non-negative real number to each closed subspace. If $H_i$ is a collection of mutually orthogonal subspaces with closed linear span $S$ (Gleason, 1957), then:

$$\mu(S) = \sum_i \mu(A_i) \tag{B.2}$$

where the function $\mu$ can be obtained by selecting a vector $v$, and for each closed subspace $H$, taking $\mu(H)$ as the square of the norm of the projection of $v$ on $H$.

For a separable Hilbert space $H$ with dimension at least 3, function $\mu$ is a measure on the closed subspace, then there is a positive self-adjoint operator $M$:

$$\mu(H) = tr(MP_H) \tag{B.3}$$

where $P_H$ is the orthogonal projection on $H$, which defines a measure on the closed subspaces. $M$ denotes the density operator.

The theorem gives a general method for measuring events in quantum logic computations, that is, events in Hilbert spaces can be represented by a self-adjoint operator (named density operator), and assures that the intuitive notion of quantum state is perfectly grasped by the notion of density operator (Dalla Chiara *et al.*, 2007).

## B.2 Quantum Interference

This experimental phenomenon strongly proves the wave theory (that is, light has the characteristics of waves, so the nature of light is a wave). This experiment has aroused wide attention in the physics community. Albert Einstein proved that photons have wave-particle duality in the early 20th century (Yang, 2005). The French physicist De Broglie further promoted this theory and proposed the matter wave theory (De Broglie, 1923), which calculates the probability that a particle will appear at a certain point in space by studying its fluctuations.

The process of quantum interference can be formalized by linear algebra operations. Suppose $|\psi_1\rangle$ and $|\psi_2\rangle$ are the quantum states of particle passing through two slits $A$ and $B$ to reach the shield

respectively. According to the principle of quantum superposition, the quantum state measured by the shield is:

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle \tag{B.4}$$

where $c_1$ and $c_2$ satisfy $|c_1|^2 + |c_2|^2 = 1$, and the probability of particles reaching the shield is:

$$\langle\psi|\psi\rangle = |c_1|^2 \langle\psi_1|\psi_1\rangle + |c_2|^2 \langle\psi_2|\psi_2\rangle + c_1 c_2^* \langle\psi_2|\psi_1\rangle + c_1 c_2^* \langle\psi_1|\psi_2\rangle \tag{B.5}$$

where $c_1 c_2^* \langle\psi_2|\psi_1\rangle + c_1 c_2^* \langle\psi_1|\psi_2\rangle$ is the quantum interference term, which represents the interference effect between two beams of particles.

After Young's double-slit experiment, researchers conducted a series of variant experiments and observed more evidence of quantum theory.

In the interference experiment of single particle, the researchers ensure that the source only emitted one photon at a time, and simultaneously open the slit $A$ and slit $B$, the light and dark stripes can be seen on the shield, resulting in the phenomenon of quantum interference. This further proves that interference not only exists between particles, but particles can also interfere with themselves. This experimental phenomenon proves the principle of quantum superposition state, that is, in the interference experiment this photon passes through both slit $A$ and slit $B$. These experiments led researchers to wonder, which slit did the photon pass through? Therefore, the researchers placed detectors at the two slits to try to detect the path of photons. However, the experimental results found that after adding the detector, the interference fringes disappeared. The reason was that the observations caused the collapse of the superposition state, thereby destroying the interference phenomenon. Quantum measurement is reflected in this experiment as a bridge from micro to macro.

# C

# Quantum Many-body Problem and Neural Network

Neural network and quantum many-body theory are faced with similar problems and objects, that is to extract key features from a large number of information (Zi, 2017). Neural network needs to extract feature information from images or texts, and quantum many-body theory needs to obtain the required particle characteristics from the complex quantum system. Meanwhile, neural networks have strong learning and representation ability, and quantum many-body has a profound theoretical foundation, which inspires researchers to conduct cross research on neural networks and quantum many-body.

## C.1 The Challenges of Quantum Many-body Theory

The wave function is an effective tool to represent the quantum state, and its information is proportional to the complexity of the quantum state. Therefore, the sufficient coding of many-body quantum state requires an exponential amount of information. However, in the case of fact, the wave function describing many physical many-body systems only needs limited information, so many researchers try to use a limited number of quantum entanglement and a small number of physical states to solve the representation of many-body quantum states (Verstraete *et al.*, 2006).

Currently, there are mainly two kinds of methods to solve the quantum many-body problem. The first method is based on the numerical approximation of wave function, such as the quantum Monte Carlo method (Ceperley and Alder, 1986, Carlson *et al.*, 2015). The second method is to compress the quantum state effectively. Matrix product state (White, 1992, Schollwoeck, 2011) and tensor network (Orus, 2014, Verstraete *et al.*, 2008) are both effective compression tools.

The above two types of solutions to quantum many-body problems are still not considered a universal solution, such as dynamics of high-dimensional systems, precise ground state properties of strongly interacting fermions, *etc.*, have not been solved.

## C.2 From Neural Network to Quantum Many-body

The Carleo and Troyer (2016) solution uses machine learning to simulate quantum many-body wave functions and shows excellent performance beyond numerical simulation methods. Given the quantum many-body system $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N)$ solution, we hope to obtain the many-body wave function which can represent the amplitude and phase of $\mathcal{S}$. The core idea of Neural-Network Quantum States (NQS) is to use the $\Psi(s)$ trained by neural network to approximate the system $\mathcal{S}$. Here we discuss the spin 1/2 quantum system by restricted Boltzmann machine (RBM). Among them, RBM consists of N visible layers, corresponding to the many-body configuration $\mathcal{S}$; the input is multiplied by a weight matrix $W_{ij}$ obtains M hidden layers corresponding to auxiliary spin variables; and finally generates output through non-linear activation function. The process is shown in Figure C.1.

The energy function of RMB is as follows:

$$E = -\sum_i a_i v_i - \sum_j b_j h_j - \sum w_{ij} v_i h_j \tag{C.1}$$

According to the Hammersley-Clifford Theorem, the quantum state $\mathcal{S}$ can be expressed as:

$$\Psi_M(\mathcal{S}; \mathcal{W}) = \sum_{h_i} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z} \tag{C.2}$$

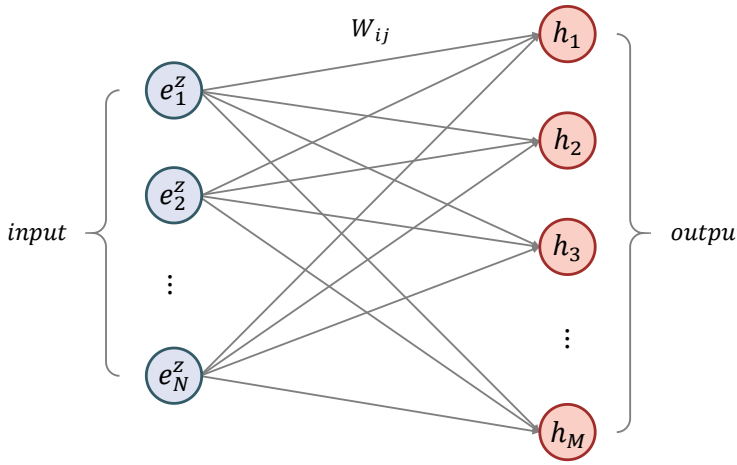



**Figure C.1:** Nerual Network Encoding Many-body System.

where $h_i = \{-1, 1\}$, $\mathcal{S} = (\sigma_1^z, \ldots, \sigma_N^z)$, $a_j$, $b_i$ and $W_{ij}$ are parameters of neural network. In order to update network parameters, Equation C.2 can be changed to:

$$\Psi_M(\mathcal{S}; \mathcal{W}) = e^{\sum_j a_j \sigma_j^z} \times \prod_{i=1}^{M} 2cosh[b_i + \sum_j W ij \sigma_j^z] \tag{C.3}$$

However, in the practical application, the state of the many-body system is unknown, and the supervised learning method cannot be used to solve $\Psi$. Therefore, a reinforcement learning method based on the variational principle is proposed. For a given Hamiltonian H, optimal representation of the unknown ground state of H is solved by minimizing the expectation value of the energy relative to the network weight W. The expectation value of the energy can be expressed as:

$$E(\mathcal{W} =< \Psi_M \mid \mathcal{H} \mid \Psi_M > / < \Psi_M \mid \Psi_M > \tag{C.4}$$

There are also some novel works around the study of quantum many-body theory based on neural networks:

- Gao and Duan (2017) reveals the close relationship between deep neural networks and quantum many-body problems. This work

first proves that any shallow neural network can not effectively represent the general quantum many-body wave function. Next, it is proved that the deep neural network can effectively represent the quantum states generated by any quantum dynamics evolution, indicating that the number of required parameters only increases linearly with the number of particles and the evolution time. In addition, this study also proves that the ground state of any common physical system can be effectively represented by deep neural network, and a deep learning algorithm is proposed to calculate the wave function generated by the ground state or quantum dynamics evolution.

- Using neural network to update the quantum many-body algorithm, or propose a new algorithm. For example, using the idea of machine learning to optimize the quantum Monte Carlo method (Huang and Wang, 2017; Liu *et al.*, 2017) and propose the cluster updating algorithm (Wang, 2017).

## C.3 From Quantum Many-body to Neural Network

The methods and ideas of the quantum many-body problem can also promote neural network. The problem with neural network is its poor interpretability and lack of strong theory to support the design and optimization of neural network structure. The related work involves the interpretation of neural networks from the perspective of physics (Mehta and Schwab, 2014; Lin *et al.*, 2017; Levine *et al.*, 2017; Levine *et al.*, 2017), and the application of quantum many-body method in machine learning (Stoudenmire and Schwab, 2016). In the field of neural language modeling, language modeling techniques based on quantum many-body have also been proposed (Section 3.3).

# D

## Variants of Quantum Language Model

Many researchers focus on the further development of QLM, and have derived many enlightening works which involve the application of quantum theoretical properties in language modeling. This section introduces the following work:

- Based on the equivalence between quantum entanglement and statistical Unconditional Pure Dependence (UPD) in post-measurement configuration (Hou and Song, 2009; Hou *et al.*, 2013), Xie *et al.* (2015) proposed explicitly modeling quantum entanglement in text by extracting UPD pattern.
- Zhang *et al.* (2018c) proposed a Golobal-convergence based Quantum Language Model (GQLM) unsupervised sentiment analysis approach, which proved the feasibility of quantum theory in sentiment analysis. Specifically, this work constructs density matrices for dictionaries and documents, and employs the quantum relative entropy to judge the similarity between dictionaries and documents.
- Most of the existing QLM models only query words without considering query extensions in IR. Based on GQLM (Zhang *et*

*al.*, 2018c), Li *et al.* (2018a) proposed the Quantum Language Model-based Query Expansion (QLM-QE) framework in order to take more terms into modeling.

- In Blacoe *et al.* (2013), a novel semantic space model was proposed to capture lexical meaning. Compared with Sordoni *et al.* (2013), this model takes word order into account and quantum correlation properties are also considered.
- Blacoe (2014) develops the Blacoe *et al.* (2013) method (*i.e.*, learning density operators from dependency-parsed corpus) and extends their model to complete sentences. In addition, the entanglement between syntactic relations in linguistic density operators is analyzed.
- Basile and Tamburini (2017) proposed a QLM based on the unitary evolution of a quantum state in time and applied to speech recognition. It is worth noting that an ancillary system is proposed in this work to avoid the disappearance of quantum effects at complete collapse.

# E

# Tensor Networks and Interpretability

## E.1 Fundamental Connection between Tensor Network and Neural Network

Based on the language representation of Section 3.3, Zhang *et al.* (2018b) and Zhang *et al.* (2019a) further study the fundamental relationship between tensor network and neural network in language modeling. Specifically, the high-dimensional tensor decomposed by tensor decomposition can derive the neural language models based on Convolutionnal Neural Network (CNN) and Recurrent Neural Network (RNN).

### E.1.1 The Representation of Tensor Network

The Tensor Network (TN) is an effective tool to solve the problem of complex quantum system (Kohn, 1999). Its high dimensional tensors and tensor decomposition are more suitable for the fitting and calculation of complex systems. In this section, we introduce the basic concepts of tensor networks and common decomposition methods.

Tensor networks can be represented by graphical representation. As shown in Figure E.1, nodes represent algebraic objects and the edges are different coordinates or indicators. Figure E.1(a) is a zero-

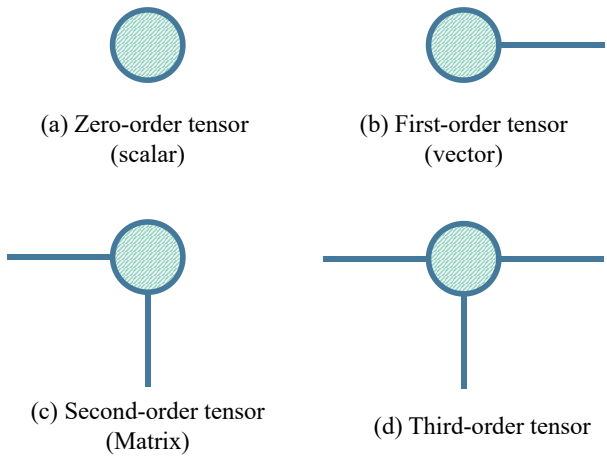



**Figure E.1:** The graphical representation of tensors.

order tensor, which can be considered as a scalar, Figure E.1(b) is a vector, which represents first-order tensor, Figure E.1(c) is a matrix and Figure E.1(d) represents a third-order tensor. The probability amplitude of the quantum many-body wave function of Equation 3.17 can be expressed as a $N$-order tensor (Figure E.2).

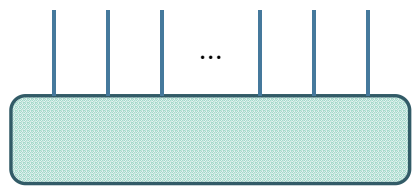



**Figure E.2:** A N-order tensor.

### Tensor Product and Tensor Contraction

Tensor product (represented as $\otimes$) is a basic operation in tensor networks, which maps two low-order tensors to a higher-order tensor and expands two vector spaces into a high-dimensional vector space. For example,

the tensor product of two first-order tensors $\alpha_i \in \mathbb{R}^m$ and $\alpha_j \in \mathbb{R}^d$ can be obtained, where $\alpha_i \otimes \alpha_j \in \mathbb{R}^{m \times m}$:

$$\alpha_i \otimes \alpha_j = \begin{bmatrix} \alpha_{i,1}\alpha_{j,1} & \cdots & \alpha_{i,m}\alpha_{j,1} \\ \vdots & \ddots & \vdots \\ \alpha_{i,1}\alpha_{j,m} & \cdots & \alpha_{i,m}\alpha_{j,m} \end{bmatrix} \tag{E.1}$$

Here we can define the rank-one tensor, that is, if a $n$-order tensor $\mathcal{A}$ can be written in the following form, then the tensor $\mathcal{A}$ is a rank-one tensor:

$$\mathcal{A} = \alpha_1 \otimes \alpha_2 \cdots \otimes \alpha_n \tag{E.2}$$

If tensor $\mathcal{B}$ can be decomposed into the sum of $n$ rank-one tensors, then the rank of this tensor $\mathcal{B}$ is $n$.

There are two third-order tensor $\mathcal{A}_{smj}$ and $\mathcal{B}_{jpq}$ shown as Figure E.1(d). They have a same indicator $j$ with $d$ dimensions, then the contraction between $\mathcal{A}_{smj}$ and $\mathcal{B}_{jpq}$ is:

$$\mathcal{D}_{smpq} = \sum_{=1}^{d} \mathcal{A}_{smj}\mathcal{B}_{jpq} \tag{E.3}$$

where the tensor $\mathcal{D}_{smpq}$ after tensor contraction is shown as Figure E.3.

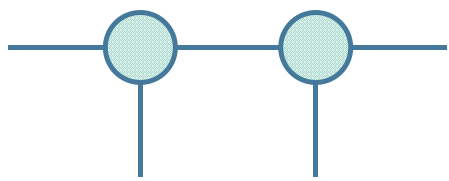


**Figure E.3:** Tensor contraction.

### Tensor Network States

At present, the common TN states used in *Machine Learning (ML)* are *Matrix Product State (MPS)* (Liu *et al.*, 2017), which is shown in Figure E.4, and *Tree Tensor Network state (TTN)* (Carrasquilla and Melko, 2017) demonstrated in Figure E.5. The legs connecting two objects are virtual indicators, which can be seen as the rank after decomposition.

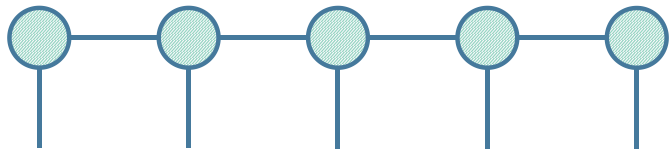


**Figure E.4:** Matrix Product State (MPS).

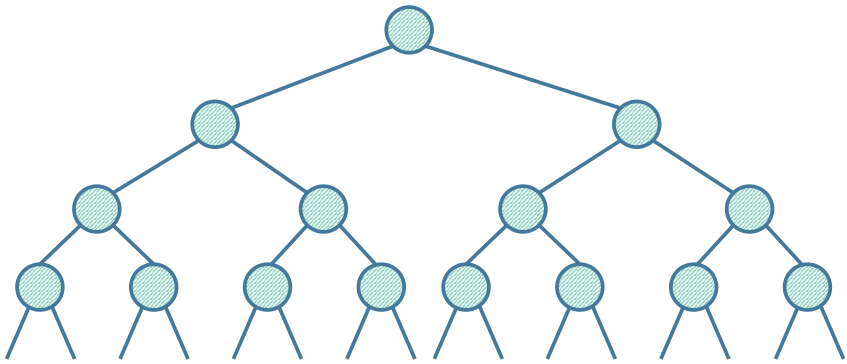


**Figure E.5:** Tree Tensor Network state (TTN).

The physical indicators of tensor networks are represented by the open legs connecting one object in Figure E.4 and Figure E.5.

The initial TN is random, so how to make the TN gradually and correctly shrink to the state of approximate wave function is a problem that must be solved. The TN algorithm to solve these problems and adapt to ML is the variational algorithm, including Density Matrix Renormalization Group algorithm (DMRG) (Wang, 2016), variational PEPS (Bradde and Bialek, 2017), variational MERA (Carleo and Troyer, 2016) and Tensor Network Renormalization group algorithm (TNR) (Bény, 2013).

### E.2 Tensor Network and Machine Learning

This section proposes why *Tensor Network (TN)* can be used for *Machine Learning (ML)*. One is that the ML and TN have some common points. The TN states can effectively approximate the main information of quantum many-body wave function. It is worth noting that the ML (e.g., deep learning) models use some parameters to fit unknown and complex functions (Glasser *et al.*, 2019). Both of them use limited parameters to fit or approximate an exponential function space according to the prior knowledge of data. On the other hand, TNs are the extension of probability graph models, which have similar properties with some classic probability models (e.g., Hidden Markov Model). Moreover, end-to-end learning of deep neural network also inspires the development of TNs, and TNs take theoretical interpretability for neural network (Khrulkov *et al.*, 2018; Levine *et al.*, 2019).

Currently, TN and ML have established close connections. Tensor network machine learning models can be divided into two kinds, which are tensor network states (Stoudenmire and Schwab, 2016; Cheng *et al.*, 2019) and mixed models between neural network and tensor network (Gao *et al.*, 2020). Similarly, tensor network machine learning algorithms can directly deploy on quantum hardwares to solve some problems in physics and mathematics.

Stoudenmire and Schwab (2016) were the first to use TN (*i.e.*, MPS) model ML tasks (*i.e.*, image recognition). In order to transfer the TN to ML image domain, the work constructed the wave function representation of the image. Let a pixel be a particle in a many-body system, then a wave function with N pixels is expressed as follows:

$$\Phi^{s_1 s_2 \cdots s_N}(x) = \Phi^{s_1}(x_1) \otimes \Phi^{s_2}(x_2) \otimes \cdots \otimes \Phi^{s_N}(x_N) \tag{E.4}$$

If the basic vector dimension of a pixel is $d(d > 1)$, then a picture is represented by a high-order tensor of $d^N$ dimension. But in an image, a pixel has a single value, that is, its basis vector dimension is 1. Therefore, researchers need to construct the superposition state representation of the pixel to map to the high-dimensional space. The mapping of a single pixel is called local feature mapping. There are many ways of feature

mapping and the mapping used in Stoudenmire and Schwab (2016) is:

$$\Phi^{s_j}(x_j) = [\cos(\tfrac{\pi}{2}x_j\ ), \sin(\tfrac{\pi}{2}x_j\ )]$$

This kind of local mapping satisfies that the sum of squares of probability amplitude is 1, but it does not need to be so strict in practice.

As shown in Figure E.6, suppose that the data has $\ell$ tags, then the output of the model, also known as the decision function of the model, can be defined as:

$$f^{\ell}(\mathbf{x}) = W^{\ell} \cdot \Phi(\mathbf{x}) \tag{E.5}$$

where $f^{\ell}$ is a $\ell$-dimensional vector. they use $W^{\ell}$ to represent an abstract supervised model, which is a mapping from a $d^N$ dimensional vector to an $\ell$-dimensional vector.

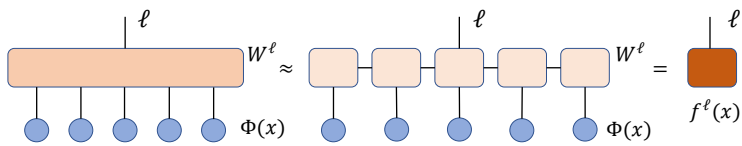



**Figure E.6:** The decision function $f^{\ell}$ $(\mathbf{x})$ is defined by the contraction of $W^{\ell}$ and specific input $\Phi(\mathbf{x})$ . The label $\ell$ of maximum value of $f^{\ell}$ $(\mathbf{x})$ is the label of $\mathbf{x}$ predicted by the model.

If $W^{\ell}$ is represented by a higher-order tensor, the size of the parameter tensor is $N_L \cdot d^N$, where $N_L$ is the number of tags, then it is unrealistic to calculate this tensor strictly.

Therefore, an efficient approximation method is needed to optimize the tensor. This method is to approximate a higher-order tensor by a tensor network composed of a set of low-order tensors, which can be approximated by a MPS.

$$W^{\ell}_{s_1 s_2 \cdots s_N} = \sum_{\{a\}} A^{a_1}_{s_1} A^{a_1 a_2}_{s_2} \cdots A^{\ell; a_j a_{j+1}}_{s_j} \cdots A^{a_{N-1}}_{s_N} \tag{E.6}$$

For this model, the dimension of the output vector $\ell$ is equal to the number of tags. For MPS, its output position is not fixed. In actual training, it is usually output at $\frac{1}{2}$ of MPS length.

The loss function proposed is Cross Entropy, and the MPS optimization algorithm is a DMRG like algorithm. MPS is optimized and updated by DMRG like algorithm.

$$\mathrm{C} = \tfrac{1}{2} \sum_{n=1}^{N_T} \sum_{\ell} f^{\ell}(x_n - \delta^{\ell}_{L_n})^2$$

In the actual training process, they chosen to control the size of MPS by setting the bond dimension of hyper-parameter, and use the back propagation method of neural network to train.

### E.3 Matrix Product Operators for Sequence to Sequence Learning

In physics, MPS are used to represent wave-functions, probability distributions or density matrices as a product of tensors. As an extension, *Matrix Product Operators (MPO)* can represent quantum mechanical operators that map an MPS to another one. Since MPS can give an accurate description of correlations between distant parts of a system (like a sequence), MPO can be used to do sequence to sequence prediction. Guo *et al.* (2018) trains an MPO which can accurately reproduce the transformation from input vectors to output vectors, and apply to sequence to sequence prediction.

The MPO here is able to generate a sequence of $L$ labels given another one of the same length. If it is stated as mapping between vector space $\mathcal{X}$ and $\mathcal{Y}$ that is:

$$\vec{y}_i = f\left(\vec{x}_i\right) \tag{E.7}$$

where $\vec{x}_i$ and $\vec{y}_i$ is vector in $\mathcal{X}$ and $\mathcal{Y}$. The MPO would try to provide an accurate approximation $\tilde{\vec{y}}_i$ to the exact vector $\vec{y}_i$.

The probability of a given sequence of integer numbers (*i.e.*, $P(\vec{\sigma})$) can be written as:

$$P(\vec{\sigma}) = \sum_{a_0,\ldots,a_L} M^{\sigma_1}_{a_0,a_1} M^{\sigma_2}_{a_1,a_2} \ldots M^{\sigma_L}_{a_{L-1},a_L} \tag{E.8}$$

where $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_L)$, $\sigma_l$ describes local degree of freedom with total number is $d$, and $a_l$ is auxiliary degree of freedom which takes correlations between different states into account. $a_l$ is also known as "bond dimension" $D$, the larger the more accurate is the probability distribution. So, input and output sequence can both be fit into MPS,

then an MPO $W_{\vec{\sigma}}^{\vec{\tau}}$ is applied, where $\tau$ is a sequence of output integers. The MPOs can be parameterized by a product of 4-dimensional tensors:

$$W_{\vec{\sigma}}^{\vec{\tau}} = \sum_{b_0,b_1,\ldots,b_L} W_{b_0,b_1}^{\sigma_1,\tau_1} W_{b_1,b_2}^{\sigma_2,\tau_2} \ldots W_{b_{L-1},b_L}^{\sigma_L,\tau_L} \tag{E.9}$$

The first step to train an MPO is to convert sequence to MPS. The input and the output sequences may belong tp different spaces. Each of input sequence item can be reprensented as a vector of finite size $d$ and $d'$ for output. In this way, each input sequence $\vec{x}_i$ can be mapped into a product MPS of physical dimension $d$:

$$\vec{x}_i \to X_i^{\vec{\sigma}} = \sum_{a_0,\ldots,a_L} X_{i,a_0,a_1}^{\sigma_1} X_{i,a_1,a_2}^{\sigma_2} \ldots X_{i,a_{L-1},a_L}^{\sigma_L} \tag{E.10}$$

noted that the different $i$ index of $X_{i,a_l,a_{l+1}}^{\sigma_l}$ corresponds to different sequences. Similarly, each output $\vec{y}_i$ can be mapped into a MPS:

$$\vec{y}_i \to Y_i^{\vec{\tau}} = \sum_{c_0,\ldots,c_L} Y_{i,c_0,c_1}^{\tau_1} Y_{i,c_1,c_2}^{\tau_2} \ldots Y_{i,c_{L-1},c_L}^{\tau_L} \tag{E.11}$$

Thus, multiplying an MPS by an MPO can produce another MPS:

$$\bar{Y}_i^{\vec{\tau}} = W_{\vec{\sigma}}^{\vec{\tau}} X_i^{\vec{\sigma}} = \sum_{\bar{c}_0,\ldots,\bar{c}_L} \bar{Y}_{i,\bar{c}_0,\bar{c}_1}^{\tau_1} \ldots \bar{Y}_{i,\bar{c}_{L-1},\bar{c}_L}^{\tau_L} \tag{E.12}$$

where $\bar{c}_l$ is a new auxiliary index given by $\bar{c}_l = (a_l, b_l)$ on site $l$, and the MPS of site $l$ is:

$$\bar{Y}_{i,\bar{c}_{l-1},\bar{c}_l}^{\tau_l} = \sum_{\sigma_l} W_{b_{l-1},b_l}^{\sigma_l,\tau_l} X_{i,a_{l-1},a_l}^{\sigma_l} \tag{E.13}$$

A cost function $C(W_{\vec{\sigma}}^{\tau})$ can now be defined as:

$$\begin{aligned} C(W\vec{\sigma}) = \sum_{i=1}^{N} \left(\bar{Y}_i^{\vec{\tau}\dagger} - Y_i^{\vec{\tau}\dagger}\right)\left(\bar{Y}_i^{\vec{\tau}} - Y_i^{\vec{\tau}}\right) \\ + \alpha \operatorname{tr}\left(W_{\vec{\sigma}}^{\vec{\tau}\dagger} W_{\vec{\sigma}}^{\vec{\tau}}\right) \end{aligned} \tag{E.14}$$

where the last term with the coefficient $\alpha$ regularizes the MPO. Moreover, an iterative approach are used that transform a global minimization problem into many local minimization problems via an iterative procedure.

$$\frac{\partial C(\hat{W})}{\partial W_{b_l-1}^{\sigma_l,\tau_l} b_l} = 0 \tag{E.15}$$

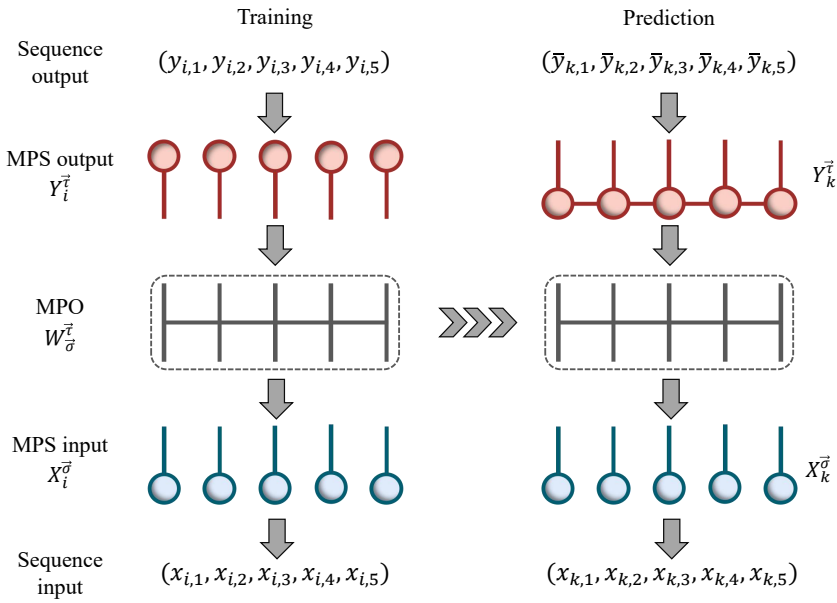



**Figure E.7:** Training phase and prediction phase of MPO for sequence learning. The trainable MPO $W_{\vec{\sigma}}^{\vec{\tau}}$ is used in prediction phase to transform $X_k^{\vec{\sigma}}$ to $\bar{Y}_k^{\vec{\tau}}$.

The minimization procedure is taken place from site $l = 1$ to $l = L$ and backward which is referred, in many body physics, as sweep. A linear solver can be used to compute the local optimal tensor $W_{b_{l-1},b_l}^{\sigma_l,\tau_l}$. The sweeps are repeated until a maximum $k_max$ or until the cost function converges to a previously determined parameter $\epsilon_t$.

As shown in Figure E.7, once the MPO has been trained, it is possible to use it to make predictions. In order to do so, first the input sequence is converted into an input MPS. This MPS is then multiplied with the trained MPO, and this results in the output matrix product state $\bar{Y}_i^{\vec{r}}$. First, approximate the output MPS with an MPS of bond dimension $D = 1$. Then, the MPS with bond dimension $D = 1$ is converted to a sequence by reversing the way in which, in precedence, a sequence $\vec{x_i}$ (or $\vec{y_i}$).

## E.4 Uniform MPS for Probabilistic Sequence Modeling

In contrast to previous models built on the fixed-size MPS tensor network where the model is parameterized by a fixed number of core tensors, the uniform matrix product state (u-MPS) model (Miller *et al.*, 2021) is parameterized by a single core, making itself recurrent in nature and able to evaluate and sample strings of arbitrary length. Specifically, the u-MPS is a recurrent-style factorization parameterized by a single core tensor $\mathcal{A}$ of shape $(D, d, D)$, along with a pair of D-dimensional boundary vectors $\alpha$ and $\omega$. Given these components, one can generate $n$th order tensors $\mathcal{T}_n \in \mathbb{K}_{d^n}$. Particularly, letting $\mathcal{A}(s_j)$ denote the $D \times D$ transition matrix obtained by fixing the middle index of $\mathcal{A}$ to $s_j \in [d]$, then the elements of $\mathcal{T}_n$ are given by:

$$\mathcal{T}_n\left(s_1, s_2, \ldots, s_n\right) = \alpha^{\dagger} \mathcal{A}\left(s_1\right) \mathcal{A}\left(s_2\right) \cdots \mathcal{A}\left(s_n\right) \omega. \tag{E.16}$$

Given the notation $\mathcal{A}(s) := \mathcal{A}(s_1)\mathcal{A}(s_2)...\mathcal{A}(s_n)$ to indicate the transition matrix appearing in Equation E.16, thus it can be thought as the u-MPS learned representation of an arbitrary string of length $n$: $s = s_1 s_2 ... s_n$. In NLP tasks, $\mathcal{A}(s)$ can be used for downstream tasks, where compositional representation of text which has arbitrary length are needed.

u-MPS is suitable to model functions of sequences because tensor $\mathbf{T}_n$ can represent outputs of the model on any sequence of $n$ symbols from an alphabet set $\Sigma = [d]$. To take an approach inspired by quantum mechanics which ties deeper with tensor network, Miller *et al.* (2021) chooses to construct with Born machine architecture. That leads to map an $n$th order tensor $\mathcal{T}_n$ to a probability distribution over length-n sequences via:

$$P\left(s_1, s_2, \ldots, s_n\right) = \left|\mathcal{T}_n\left(s_1, s_2, \ldots, s_n\right)\right|^2 / \mathcal{Z}_n. \tag{E.17}$$

This rule ensures $P\left(s_1, s_2, \ldots, s_n\right)$ sums to 1 over all sequences of length $n$. As shown in Figure E.8, although computing the normalization factor $\mathcal{Z}_n$ is intractable for arbitrary tensor $\mathcal{T}_n$, it can be exactly and efficiently computed when it is parameterized by an MPS.

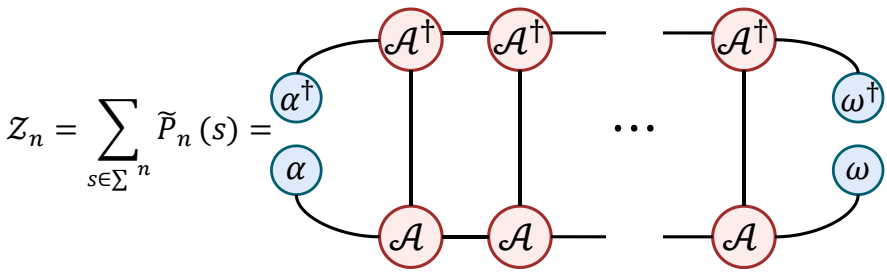



**Figure E.8:** The normalization factor $\mathcal{Z}_n$ is given by the sum over all unnormalized probabilities $\tilde{P}_n(s_1, s_2, ..., s_n)$, which can be expressed as a tensor network diagram.

If u-MPS is used in place of a finite-size MPS, equation can be modified as:

$$\begin{aligned} P_n(s) &= \tilde{P}_n(s)/\mathcal{Z}_n = \left|\alpha^\dagger \mathcal{A}(s)\omega\right|^2 /\mathcal{Z}_n \\ &= \left|\alpha^\dagger \mathcal{A}(s_1)\,\mathcal{A}(s_2)\cdots\mathcal{A}(s_n)\,\omega\right|^2 /\mathcal{Z}_n \end{aligned} . \tag{E.18}$$

That means sentence $s$ is represented of a recurrent tensor $\mathcal{A}(s)$, and the normalization factor $\mathcal{Z}_n = \Sigma_{s\in\Sigma^n}\tilde{P}_n(s)$ can be computed exactly.

## E.5 Tensor Networks and Interpretability

### E.5.1 The Background of Interpretability

Faced with practical problems, most models based on ML use a large number of input-output data learning algorithms, and then use the learned algorithms to predict the output of unlearned data. Although this learning paradigm realizes the end-to-end training, there is no doubt that it also hides a lot of learning details, which hinders the performance improvement and application feasibility of the model. Lipton (2016) proposes that the study of interpretability is helpful to:

- Reveal the causality in observational data.
- Enhance the Transferability of the model.
- Provide more useful information for users.

Specifically, the interpretability of current models mainly includes two aspects: transparency and post hoc interpretability.

- *Transparency.* How does the model work? This part of the research focuses on the self-explainability of components through some mechanisms in the designing phase of the model, enriching the understanding of the working mechanism of the model in different stages and tasks.
- *Post hoc interpretability.* What else can the model tell me? For some opaque models, the prediction results can be analyzed by some common post hoc interpretation approaches (for example, natural language explanations, visualizations of learned representations or models, and explanations by example), so as to provide useful information for users.

### E.5.2 Interpretability Research Based on Tensor Networks

It is worth noting that the research on the interpretability based on TN belongs to the field of exploring model architecture (*i.e.*, transparency), which derives from the theoretical interpretability of TN, which is mainly reflected in the following:

- TN is based on matrix decomposition algorithm (Golub and Reinsch, 2007), which makes TN discard unimportant data in principle when data compression.
- TN is a probability model, and its optimization method is linear calculation, which is more interpretable than neural network.
- In TN, quantum entanglement entropy can be used to calculate the expression ability of TNs, which means that it is of theoretical significance to adjust the structure of TNs, such as the hierarchy or dimension of TN.

CNN (Kalchbrenner *et al.*, 2014) and RNN (Schuster and Paliwal, 1997) have been successfully applied in language modeling and are surprisingly efficient at solving practical tasks, but the theory behind this

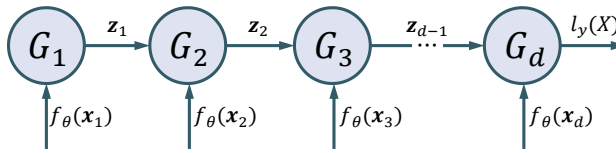



**Figure E.9:** Recurrent neural architecture corresponding to the Tensor Train decomposition.

phenomenon is only starting to catch up with the practice. Specifically, how to obtain the inductive bias of the network through the data prior, so as to effectively guide the network structure design, is one of the important issues that researchers pay attention to. The second is how to quantify the expressive ability of neural networks, and compare different network architectures through expressive ability, so as to choose a better network architecture.

In some works (Cohen and Shashua, 2016a), TNs are used as effective algorithms for measuring the expressive ability of different neural networks. Deep convolutional networks corresponding to the *Hierarchical Tucker (HT)* tensor decomposition has been proven to have exponentially higher expressive power than shallow networks corresponding to the CP-Networks. Then, Khrulkov *et al.* (2018) proved that the expressive power for RNNs corresponding to the *Tensor Train (TT)* decomposition shown in Figure E.9 is higher than shallow convolutional networks. Finally, based on the theoretical analysis of CP-Networks, HT-Networks and TT-Networks, the work systematically compared the expressive powers of deep CNNs, shallow networks and RNNs. The comparison results of expressive power is shown in Table E.1.

**Table E.1:** Comparison of the expressive power of various networks. Given a network of width $r$ (the rank of tensor decomposition), specified in a column, rows correspond to the upper bound on the width of the equivalent network of other type.

| | **TT-Network** | **HT-Network** | **CP-Network** |
|---|---|---|---|
| TT-Network | $r$ | $r^{log_2(d)}/2$ | $r$ |
| HT-Network | $r^2$ | $r$ | $r$ |
| CP-Network | $\geq r^{\frac{d}{2}}$ | $\geq r^{\frac{d}{2}}$ | $r$ |

Deep CNN models display exciting effects for various machine learning tasks. But it still has some problems, such as the hyper-parameters setting about convolution channels. Levine *et al.* (2017) established the basic connection of quantum information and deep CNN to prove the setting of channels in each convolution layer satisfying their effects in overall inductive bias which reflects network's prior knowledge or expressive ability. Based quantum entanglement entropy and graph theory, the inductive bias of the neural network can be calculated. That is, the channels of each layer could be sure.

First, they proved the equivalence between quantum many-body wave function and a deep *Convolutional Arithmetic Circuit (ConvAC)*. The projection of the many-body quantum state $\Psi$ onto the product state $|\phi\rangle$ is written as:

$$\begin{aligned}\langle \Psi^{ps}|\Psi\rangle &= \sum_{d_1,\ldots,d_N=1}^{M} \mathcal{A}^{y}_{d_1\ldots d_N} \prod_{j=1}^{N} f_{\theta_{d_j}(x_j)} \\ &= \sum_{d_1,\ldots,d_N=1}^{M} \mathcal{A}^{y}_{d_1\ldots d_N} \mathcal{A}^{ps}_{d_1\ldots d_N}(x_1,\ldots,x_N)\end{aligned} \tag{E.19}$$

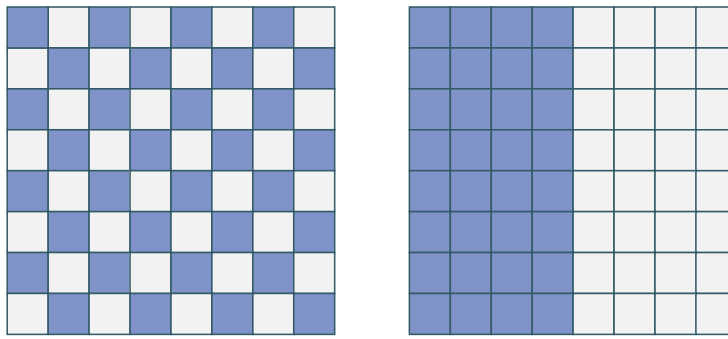
After proving the equivalence between the projection of many-body wave function and ConvAC, we will describe the measures of entanglement and correlations in Hilbert spaces, which can be used to calculate the inductive bias to measure the expressive ability of tensor networks in Levine *et al.* (2017).

There is a many-body system divided into two subsystems with entanglement, which $A = \{a_1,\ldots,a_{|A|}\}$ and $B = \{b_1,\ldots,b_{|B|}\}$. The Hilbert spaces of the system and subsystems are $\mathcal{H}, \mathcal{H}^A, \mathcal{H}^B$, respectively, with $\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$. So the wave function of the many-body system is:

$$|\Psi\rangle = \sum_{\alpha=1}^{dim(\mathcal{H}^A)} \sum_{\beta=1}^{dim(\mathcal{H}^B)} (\|\mathcal{A}\|_{A,B})_{\alpha,\beta} |\Psi^A_\alpha\rangle \otimes |\Psi^B_\beta\rangle \tag{E.20}$$

After a singular value decomposition on $\|\mathcal{A}\|_{A,B}$, the $|\Psi\rangle$ is:

$$|\Psi\rangle = \sum_{\alpha=1}^{r} \lambda_\alpha |\phi^A_\alpha \otimes |\phi^B_\beta \tag{E.21}$$



(a) Interleaved partition (b) Left-right partition

**Figure E.10:** Long-range correlations and short-range correlations.

where $\lambda_1 \geq \ldots \geq \lambda_r$ are the singular values of $\|\mathcal{A}\|_{A,B}$. So the entanglement entropy can be defined as $S = -\sum_\alpha |\lambda_\alpha|^2 ln|\lambda_\alpha|$.

Levine *et al.* (2017) proposed an example about images to explain the entanglement and correlations. As shown in Figure E.10(a), the interleaved partition and (b) is the left-right partition for an $8 \times 8$ example. These two images respectively represent the long-range correlation and short-range correlation between pixels. In order to guide ConvAC to learn different correlations more interpretably, the underlying assumption (inductive bias) is that the setting of properties in the network structure (e.g., the depth of the network and the number of channels in the network) directly affects the ability of the model to learn different correlations. First, the inductive bias is achieved by converting ConvAC into a tree-tensor network structure. Specifically, the number of channels in ConvAC can correspond to the rank of the tree tensor network, where short-range entanglement (correlation) is related to the rank of the shallow layer, and long-range entanglement (correlation) is related to the rank of the deep layer. The larger the rank, the stronger the correlation. Then, the inductive bias is used to set the convolution channel of each layer, that is, for data with different correlations, the design of the corresponding network structure is guided by the maximum flow/minimum cut theory. The specific derivation process is as follows. Then, the inductive bias is leveraged to set the convolution channel of

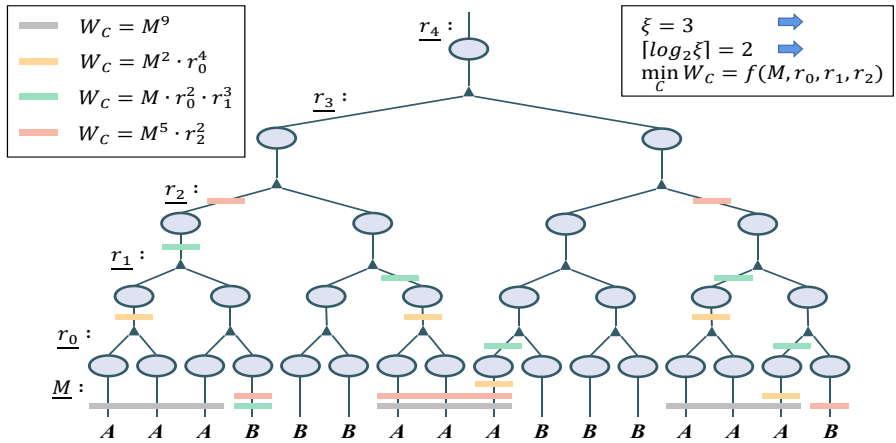



**Figure E.11:** A direct consequence of the figure is that for data characterized by short ranged correlations it is best to increase the number of channels in the lower layers, while for data characterized by long ranged correlations the channels in the deeper layers are important in order not to have 'short-cuts' harming the required expressiveness of the function realized by the network.

each layer, and according to the correlation between pixels, the design of the corresponding network structure is guided by the max-flow/min-cut algorithms. The specific derivation process is as follows.

As shown in Figure E.11, Levine *et al.* (2017) have presented a translation of the computation performed by a ConvAC to a TN. The convolutional weights are arranged as matrices (two legged nodes) placed along the network, and the same channel pooling characteristic is made available due to three legged $\delta$ tensors in a deep network, and an N + 1 legged $\delta$ tensor in a shallow network. Finally, and most importantly for their upcoming analysis, the bond dimension of each level in the TN representing the ConvAC is equal to $r_l$ , which is the number of feature maps (*i.e.*, the number of channels) comprising that level in the corresponding ConvAC architecture. So the neural network's inductive bias can be calculated, and the last question is "how to distribute the number of channels in a deep network".

Max-flow/min-cut algorithms reveal that different layers of tree-tensor networks (or ConvACs) can capture different ranges of information and correlations in data representations. For datasets with features

of feature size $D$ (*e.g.*, in a two-dimensional digit classification task, it is the size of the digit to be classified), partitions of such length scale $\varepsilon < D$ are guaranteed to be in any input position placed different parts of the elements. However, in order to correctly perform the classification task of this feature, a fine-grained function must be implemented by the network to model the strong dependencies between its different parts. As mentioned above, this means that the network must support a highly entangled metric associated with that partition, which they can then describe in terms of minimal cuts in the TN graph.

Summarizing the above arguments, for a ConvAC with pooling window size 2, if the feature size is $D$, the number of channels up to layer $l = [log_2 D]$ is more important than the number of channels in deeper layers. In natural images it may be difficult to pinpoint the most suitable feature size D, the proposed experiments demonstrate that the theory established above for deep ConvAC can better characterize the input correlations most relevant to the task at hand, and is suitable for some joins ConvAC architecture for nonlinear functions (*e.g.*, ReLU activation and pooling). Overall, the above work has two advantages in understanding deep convolutional network models:

- Through a common underlying tensor structure, theoretical connections are established between deep convolutional networks and quantum many-body wave function functions, which helps to measure the expressive power of deep networks to model complex correlation structures of their inputs using quantum entanglement. For example, leveraging entanglement entropy to model the intricate correlations structure between the two sides of an image.

- Deep convolutional networks can be constructed based on tensor networks, enabling researchers to perform graph-theoretic analysis on convolutional networks. The design satisfies the inductive bias between network expressiveness and hyper-parameters. For example, the setting of channels in each convolution layer satisfying their effects in overall inductive bias which reflects prior knowledge or expressive power of the network.

# F

## Monoidal Categories and Diagrams

In the work of Coecke *et al.* (Coecke, 2007; Clark *et al.*, 2008; Coecke *et al.*, 2010; Zeng and Coecke, 2016; Coecke *et al.*, 2020), the monoidal categories represents an intuitive operational interpretation and a purely diagrams. A monoidal category $\boldsymbol{C}$ repquires to obey the following axioms:

- A family $|\boldsymbol{C}|$ of objects:
  - for each ordered pair of objects $(A, B)$ a corresponding set $\boldsymbol{C}(A, B)$ of morphisms; it is convenient to abbreviate $f \in \boldsymbol{C}(A, B)$ by $f : A \to B$;
  - for each ordered triple of objects $(A, B, C)$, and each $f : A \to B$ and $g : B \to C$, there is a sequential composite $g \circ f : A \to C$; we moreover require that:

$$(h \circ g) \circ f = h \circ (g \circ f) \tag{F.1}$$

  - for each object $A$ there is an identity morphism $1_A : A \to A$; for $f : A \to B$ we moreover require that:

$$f \circ 1_A = f \qquad \textit{and} \qquad 1_B \circ f = f \tag{F.2}$$

- for each ordered pair of objects $(A, B)$ a composite object $A \otimes B$; we moreover require that:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) \tag{F.3}$$

- A unit object $I$ satisfies:

$$I \otimes A = A = A \otimes I \tag{F.4}$$

- for each ordered pair of morphisms $(f : A \to C, g : B \to D)$ a parallel composite $f \otimes g : A \otimes B \to C \otimes D$; we moreover require bifunctoriality, that is:

$$(g_1 \otimes g_2) \circ (f_1 \otimes f_2) = (g_1 \circ f_1) \otimes (g_2 \circ f_2) \tag{F.5}$$

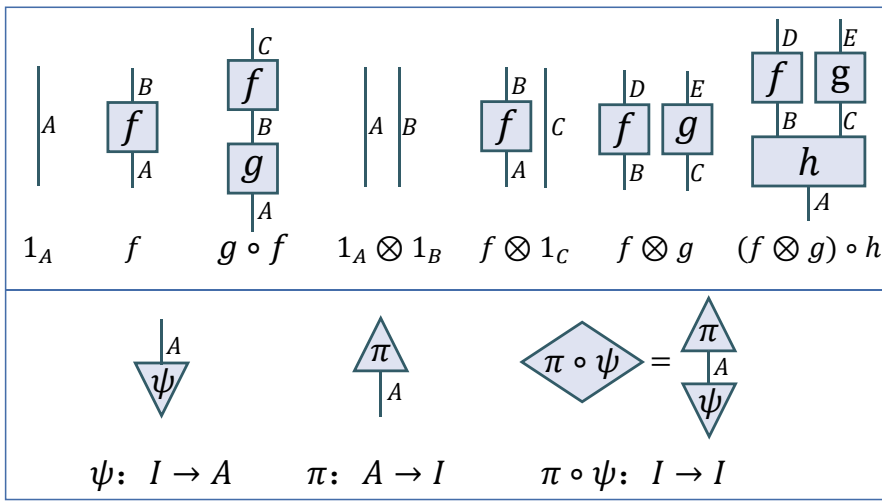



**Figure F.1:** The diagrams for monoidal categories.

If we think of the objects as types of system, and the morphism $f : A \to B$ as a process which takes a system of type $A$ as input and a system of type $B$ as output. For more information, please refer to Coecke (2006).

In the graphical calculus for monoidal categories we depict morphisms by boxes, with incoming and outgoing wires labelled by the

corresponding types, with sequential composition depicted by connecting matching outputs and inputs, and with parallel composition depicted by locating boxes side by side (Figure F.1).

# References

Aerts, D. and L. Beltran. (2022). "Are words the quanta of human language? Extending the domain of quantum cognition". *Entropy*. 24(1): 6.

Aerts, D., A. Khrennikov, M. Melucci, and B. Toni. (2019). *Quantum-Like Models for Information Retrieval and Decision-Making*. Springer.

Arora, S., Y. Liang, and T. Ma. (2017). "A Simple but Tough-to-Beat Baseline for Sentence Embeddings". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Balkır, E. (2014). "Using density matrices in a compositional distributional model of meaning". *Master's thesis, University of Oxford*.

Basile, I. and F. Tamburini. (2017). "Towards quantum language models". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 1840–1849.

Beer, K., D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf. (2020). "Training deep quantum neural networks". *Nature communications*. 11(1): 1–6.

Belkin, N. J., R. N. Oddy, and H. M. Brooks. (1982). "ASK for information retrieval: Part I. Background and theory". *Journal of documentation*.

Bengio, Y., R. Ducharme, P. Vincent, and C. Janvin. (2003). "A neural probabilistic language model". *Journal of Machine Learning Research*. 3(6): 1137–1155.

Bény, C. (2013). "Deep learning and the renormalization group". *arXiv: Quantum Physics*.

Blacoe, W. (2014). "Semantic Composition Inspired by Quantum Measurement". In: *Quantum Interaction*.

Blacoe, W., E. Kashefi, and M. Lapata. (2013). "A Quantum-Theoretic Approach to Distributional Semantics". In: *North American Chapter of the Association for Computational Linguistics*.

Borrelli, A. (2010). "Dirac's bra-ket notation and the notion of a quantum state". In: *Styles of Thinking in Science and Technology. Proceedings of the 3rd International Conference of the European Society for the History of Science*. 361–371.

Bradde, S. and W. Bialek. (2017). "PCA Meets RG". *Journal of Statistical Physics*. 167: 462–475.

Bruza, P. D. and R. J. Cole. (2006). "Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning". *Physics*. 9(10): 329–332.

Bruza, P., K. Kitto, D. Nelson, and C. McEvoy. (2009a). "Extracting spooky-activation-at-a-distance from considerations of entanglement". In: *International Symposium on Quantum Interaction*. Springer. 71–83.

Bruza, P., K. Kitto, D. Nelson, and C. McEvoy. (2009b). "Is there something quantum-like about the human mental lexicon?" *Journal of Mathematical Psychology*. 53(5): 362–377.

Bruza, P. D., Z. Wang, and J. R. Busemeyer. (2015). "Quantum cognition: a new theoretical approach to psychology". *Trends in Cognitive Sciences*. 19(7): 383–393.

Busemeyer, J. R. and P. D. Bruza. (2012). *Quantum Models of Cognition and Decision*. 1st. USA: Cambridge University Press.

Busso, C., M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan. (2008). "IEMOCAP: Interactive emotional dyadic motion capture database". *Language resources and evaluation*. 42(4): 335–359.

Buszkowski, W. (2001). "Lambek grammars based on pregroups". In: *International Conference on Logical Aspects of Computational Linguistics*. Springer. 95–109.

Cao, Z., T. Qin, T. Y. Liu, M. F. Tsai, and H. Li. (2007). "Learning to rank:from pairwise approach to listwise approach". In: *International Conference on Machine Learning*. 129–136.

Carleo, G. and M. Troyer. (2016). "Solving the Quantum Many-Body Problem with Artificial Neural Networks". *Ence*. 355(6325): 602–606.

Carlson, J., S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. Schmidt, and R. B. Wiringa. (2015). "Quantum Monte Carlo methods for nuclear physics". *Reviews of Modern Physics*. 87(3): 1067.

Carrasquilla, J. and R. G. Melko. (2017). "Machine learning phases of matter". *Nature Physics*. 13(5): 431–434.

Cavnar, W. B. and J. M. Trenkle. (1994). "N-gram-based text categorization". In: *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*. Vol. 161175.

Ceperley, D. and B. Alder. (1986). "Quantum monte carlo". *Science*. 231(4738): 555–560.

Chelba, C., T. Mikolov, M. Schuster, Q. Ge, T. Brants, P. Koehn, and T. Robinson. (2013). "One billion word benchmark for measuring progress in statistical language modeling". *arXiv preprint arXiv:1312.3005*.

Chen, H., F. X. Han, D. Niu, D. Liu, K. Lai, C. Wu, and Y. Xu. (2018). "Mix: Multi-channel information crossing for text matching". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 110–119.

Chen, j. (2020). "Research Situation and Development Strategy of Relevance of Information Retrieval". *Information Research*. 1(02): 1.

Chen, Y., Y. Pan, and D. Dong. (2021). "Quantum language model with entanglement embedding for question answering". *IEEE Transactions on Cybernetics*.

Cheng, S., L. Wang, T. Xiang, and P. Zhang. (2019). "Tree tensor networks for generative modeling". *Physical Review B*. 99(15): 155131.

Child, R., S. Gray, A. Radford, and I. Sutskever. (2019). "Generating long sequences with sparse transformers". *arXiv preprint arXiv:1904.10509*.

Clark, S., B. Coecke, and M. Sadrzadeh. (2008). "A compositional distributional model of meaning". In: *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*. Oxford. 133–140.

Clauser, J. F. and A. Shimony. (1978). "Bell's theorem. Experimental tests and implications". *Reports on Progress in Physics*. 41(12): 1881.

Coecke, B. (2006). "Introducing categories to the practicing physicist". In: *What is category theory*. Polimetrica. 45–74.

Coecke, B. (2007). "Automated Quantum Reasoning: Non Logic-Semi-Logic-Hyper-Logic". In: *AAAI Spring Symposium: Quantum Interaction*. 31–38.

Coecke, B., G. de Felice, K. Meichanetzidis, and A. Toumi. (2020). "Foundations for near-term quantum natural language processing". *arXiv preprint arXiv:2012.03755*.

Coecke, B., M. Sadrzadeh, and S. Clark. (2010). "Mathematical foundations for a compositional distributional model of meaning". *arXiv preprint arXiv:1003.4394*.

Cohen, N. and A. Shashua. (2016a). "Convolutional rectifier networks as generalized tensor decompositions". In: *International Conference on Machine Learning*. 955–963.

Cohen, N. and A. Shashua. (2016b). "Inductive bias of deep convolutional networks through pooling geometry". *arXiv preprint arXiv:1605.06743*.

Conneau, A., D. Kiela, H. Schwenk, L. Barrault, and A. Bordes. (2017). "Supervised learning of universal sentence representations from natural language inference data". *arXiv preprint arXiv:1705.02364*.

Craswell, N. (2017). "Neural Models for Full Text Search". In: *the Tenth ACM International Conference*.

Dai, B., J. Li, and R. Xu. (2020). "Multiple positional self-attention network for text classification". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 7610–7617.

Dai, Z., C. Xiong, J. Callan, and Z. Liu. (2018). "Convolutional neural networks for soft-matching n-grams in ad-hoc search". In: *Proceedings of the eleventh ACM international conference on web search and data mining*. ACM. 126–134.

Dai, Z., Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov. (2019). "Transformer-xl: Attentive language models beyond a fixed-length context". *arXiv preprint arXiv:1901.02860*.

Dalla Chiara, M. L., R. Giuntini, and M. Rédei. (2007). "The History of Quantum Logic". In: *The Many Valued and Nonmonotonic Turn in Logic*. Ed. by D. M. Gabbay and J. Woods. Vol. 8. *Handbook of the History of Logic*. North-Holland. 205–283.

De Broglie, L. (1923). "Waves and quanta". *Nature*. 112(2815): 540–540.

Dowty, D. R., R. Wall, and S. Peters. (2012). *Introduction to Montague semantics*. Vol. 11. Springer Science & Business Media.

Eisert, J., M. Friesdorf, and C. Gogolin. (2015). "Quantum many-body systems out of equilibrium". *Nature Physics*. 11: 124–130.

Eisert, J. (2013). "Entanglement and tensor network states". *arXiv preprint arXiv:1308.3318*.

Fagan, J. L. (1987). "Automatic phrase indexing for document retrieval". In: *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Frommholz, I., B. Piwowarski, M. Lalmas, and C. Rijsbergen. (2011). "Processing Queries in Session in a Quantum-Inspired IR Framework". In: *ECIR*.

Gan, G., P. Zhang, S. Li, X. Lu, and B. Wang. (2022). "MorphTE: Injecting Morphology in Tensorized Embeddings". In: *Advances in Neural Information Processing Systems*. Ed. by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho. URL: https://openreview.net/forum?id=7ilJhkpm1H.

Gao, Z.-F., S. Cheng, R.-Q. He, Z. Xie, H.-H. Zhao, Z.-Y. Lu, and T. Xiang. (2020). "Compressing deep neural networks by matrix product operators". *Physical Review Research*. 2(2): 023300.

Gao, J., J. Y. Nie, G. Wu, and G. Cao. (2004). "Dependence language model for information retrieval". In: *Sigir: International Acm Sigir Conference on Research & Development in Information Retrieval*.

Gao, X. and L. M. Duan. (2017). "Efficient representation of quantum many-body states with deep neural networks". *Nature Communications*. 8(1): 662.

Gkoumas, D., S. Uprety, and D. Song. (2018). "Investigating non-classical correlations between decision fused multi-modal documents". In: *International Symposium on Quantum Interaction*. Springer. 163–176.

Glasser, I., R. Sweke, N. Pancotti, J. Eisert, and I. Cirac. (2019). "Expressive power of tensor-network factorizations for probabilistic modeling". *Advances in neural information processing systems*. 32.

Gleason, A. M. (1957). "Measures on the closed subspaces of a Hilbert space". *Journal of mathematics and mechanics*: 885–893.

Golub, G. and C. Reinsch. (2007). "Singular value decomposition and least squares solutions". *Numerische Mathematik*. 14: 403–420.

Gonçalves, D. S., M. A. Gomes-Ruggiero, and C. Lavor. (2013). "Global convergence of diluted iterations in maximum-likelihood quantum tomography". *arXiv preprint arXiv:1306.3057*.

Graves, A. (2013). "Generating sequences with recurrent neural networks". *arXiv preprint arXiv:1308.0850*.

Gu, J., J. Bradbury, C. Xiong, V. O. Li, and R. Socher. (2017). "Non-autoregressive neural machine translation". *arXiv preprint arXiv: 1711.02281*.

Guo, C., Z. Jie, W. Lu, and D. Poletti. (2018). "Matrix product operators for sequence-to-sequence learning". *Physical Review E*. 98(4): 042114.

Guo, J., Y. Fan, Q. Ai, and W. B. Croft. (2016a). "A deep relevance matching model for ad-hoc retrieval". In: *Proceedings of the 25th ACM international on conference on information and knowledge management*. 55–64.

Guo, J., Y. Fan, Q. Ai, and W. B. Croft. (2016b). "A Deep Relevance Matching Model for Ad-hoc Retrieval". In: *ACM International on Conference on Information and Knowledge Management*. 55–64.

He, B. (2009). "Probability Ranking Principle". *Springer US*.

Hill, F., K. Cho, and A. Korhonen. (2016). "Learning Distributed Representations of Sentences from Unlabelled Data". In: *HLT-NAACL*.

Hitchcock, F. L. (1927). "The expression of a tensor or a polyadic as a sum of products". *Journal of Mathematics and Physics*. 6(1-4): 164–189.

Horodecki, R., P. Horodecki, M. Horodecki, and K. Horodecki. (2009). "Quantum entanglement". *Reviews of modern physics*. 81(2): 865.

Hou, Y. and D. Song. (2009). "Characterizing Pure High-Order Entanglements in Lexical Semantic Spaces via Information Geometry". In: *Quantum Interaction, Third International Symposium, QI 2009, Saarbrücken, Germany*. Vol. 5494. Springer. 237–250. DOI: 10.1007/978-3-642-00834-4_20.

Hou, Y., X. Zhao, D. Song, and W. Li. (2013). "Mining pure high-order word associations via information geometry for information retrieval". *ACM Transactions on Information Systems (TOIS)*. 31(3): 1–32.

Hrinchuk, O., V. Khrulkov, L. Mirvakhabova, E. Orlova, and I. Oseledets. (2019). "Tensorized embedding layers for efficient model compression". *arXiv preprint arXiv:1901.10787*.

Hu, B., Z. Lu, H. Li, and Q. Chen. (2015). "Convolutional Neural Network Architectures for Matching Natural Language Sentences". *Advances in neural information processing systems*. 3.

Hu, M. and B. Liu. (2004). "Mining and summarizing customer reviews". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168–177.

Huang, L. and L. Wang. (2017). "Accelerated Monte Carlo simulations with restricted Boltzmann machines". *Phys.rev.b*. 95(3): 035105.

Huertas-Rosero, A. F., L. Azzopardi, and C. J. V. Rijsbergen. (2008). "Characterising through Erasing: A Theoretical Framework for Representing Documents Inspired by Quantum Theory". *Computer Science*.

Ingwersen, P. (1992). *Information retrieval interaction*. Vol. 246. Taylor Graham London.

Ingwersen, P. (1996). "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory". *Journal of documentation*.

Jiang, Y., P. Zhang, H. Gao, and D. Song. (2020). "A Quantum Interference Inspired Neural Matching Model for Ad-Hoc Retrieval". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '20*. New York, NY, USA. 19–28. DOI: 10.1145/3397271.3401070.

Kalchbrenner, N., E. Grefenstette, and P. Blunsom. (2014). "A convolutional neural network for modelling sentences". *arXiv preprint arXiv:1404.2188*.

Karpathy, A., J. Johnson, and L. Fei-Fei. (2015). "Visualizing and understanding recurrent networks". *arXiv preprint arXiv:1506.02078*.

Kenter, T., A. Borisov, C. V. Gysel, M. Dehghani, M. D. Rijke, and B. Mitra. (2017). "Neural Networks for Information Retrieval". In: *the 40th International ACM SIGIR Conference*. 1403–1406.

Khrulkov, V., A. Novikov, and I. Oseledets. (2018). "Expressive power of recurrent neural networks". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=S1WRibb0Z.

Killoran, N., T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd. (2019). "Continuous-variable quantum neural networks". *Physical Review Research*. 1(3): 033063.

Kim, Y. (2014). "Convolutional Neural Networks for Sentence Classification". *CoRR*. abs/1408.5882. arXiv: 1408.5882.

Kiros, R., Y. Zhu, R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler. (2015). "Skip-Thought Vectors". *ArXiv*. abs/1506.06726.

Kitaev, A. and J. Preskill. (2006). "Topological entanglement entropy". *Physical review letters*. 96(11): 110404.

Kohn, W. (1999). "Nobel Lecture: Electronic structure of matter-wave functions and density functionals". *Reviews of Modern Physics*. 71: 1253–1266.

Kolda, T. G. and B. W. Bader. (2009). "Tensor decompositions and applications". *SIAM review*. 51(3): 455–500.

Krovetz, R. (1993). "Viewing morphology as an inference process". *Artificial Intelligence*. 118(1): 277–294.

Landsman, N. P. (2009). "Born rule and its interpretation". In: *Compendium of quantum physics*. Springer. 64–70.

Lee, J., E. Mansimov, and K. Cho. (2018). "Deterministic non-autoregressive neural sequence modeling by iterative refinement". *arXiv preprint arXiv: 1802.06901*.

Levine, Y., O. Sharir, N. Cohen, and A. Shashua. (2019). "Quantum Entanglement in Deep Learning Architectures". *Phys. Rev. Lett.* 122(6): 065301. DOI: 10.1103/PhysRevLett.122.065301.

Levine, Y., D. Yakira, N. Cohen, and A. Shashua. (2017). "Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design". arXiv: 1704.01552 [cs.LG].

Li, Q., M. Melucci, and P. Tiwari. (2018a). "Quantum Language Model-Based Query Expansion". In: *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval. ICTIR '18*. Tianjin, China: Association for Computing Machinery. 183–186. DOI: 10.1145/3234944.3234970.

Li, Q., S. Uprety, B. Wang, and D. Song. (2018b). "Quantum-inspired complex word embedding". *arXiv preprint arXiv:1805.11351*.

Li, Q., B. Wang, and M. Melucci. (2019). "CNM: An Interpretable Complex-valued Network for Matching". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics. 4139–4148.

Li, S., P. Zhang, Guobing, X. Lu, B. Wang, J. Wei, and X. Jiang. (2022). "Hypoformer: Hybrid Decomposition Transformer for Edge-friendly Neural Machine Translation". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*.

Li, X., D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu. (2016). "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network". In: *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE. 352–359.

Li, X. and D. Roth. (2002). "Learning question classifiers". In: *COLING 2002: The 19th International Conference on Computational Linguistics*.

Lin, H. W., M. Tegmark, and D. Rolnick. (2017). "Why Does Deep and Cheap Learning Work So Well?" *Journal of Statistical Physics*. 168(6): 1223–1247. DOI: 10.1007/s10955-017-1836-5.

Lipton, Z. (2016). "The Mythos of Model Interpretability". *Communications of the ACM*. 61(Oct.). DOI: 10.1145/3233231.

Liu, J., Y. Qi, Z. Y. Meng, and L. Fu. (2017). "Self-learning Monte Carlo method". *Physical Review B*. 95(4).

Lvovsky, A. I. (2003). "Iterative maximum-likelihood reconstruction in quantum homodyne tomography". *Journal of Optics B Quantum and Semiclassical Optics*. 6(6): 556–559.

Ma, X., P. Zhang, X. Mao, Y. Zhang, N. Duan, Y. Hou, and M. Zhou. (2020). "Interpretable Network Structure for Modeling Contextual Dependency". URL: https://openreview.net/forum?id=BkgUB1SYPS.

Ma, X., P. Zhang, S. Zhang, N. Duan, Y. Hou, M. Zhou, and D. Song. (2019). "A tensorized transformer for language modeling". *Advances in Neural Information Processing Systems*. 32: 2232–2242.

Manning, C. (2016). "Understanding Human Language: Can NLP and Deep Learning Help?" In: *the 39th International ACM SIGIR conference*.

Marcus, M., B. Santorini, and M. A. Marcinkiewicz. (1993). "Building a large annotated corpus of English: The Penn Treebank".

Marinescu, D. C. and G. M. Marinescu. (2012). "CHAPTER 2 - Measurements and Quantum Information". In: *Classical and Quantum Information*. Ed. by D. C. Marinescu and G. M. Marinescu. Boston: Academic Press. 133–220. DOI: https://doi.org/10.1016/B978-0-12-383874-2.00002-3.

McClean, J. R., S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. (2018). "Barren plateaus in quantum neural network training landscapes". *Nature communications*. 9(1): 1–6.

Mehta, P. and D. J. Schwab. (2014). "An exact mapping between the Variational Renormalization Group and Deep Learning". *Computer ence*.

Melucci, M. (2005). "Can Vector Space Bases Model Context?" In: *IN PROC. OF CONTEXT-BASED IR WORKSHOP*. CEUR.

Melucci, M. (2015). *Introduction to Information Retrieval and Quantum Mechanics*. Springer Berlin Heidelberg.

Merity, S., C. Xiong, J. Bradbury, and R. Socher. (2016). "Pointer sentinel mixture models". *arXiv preprint arXiv:1609.07843*.

Metzler, D. and W. B. Croft. (2005). "A Markov random field model for term dependencies". In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. 472–479.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean. (2013). "Distributed Representations of Words and Phrases and their Compositionality". *Advances in neural information processing systems*. 26.

Miller, J., G. Rabusseau, and J. Terilla. (2021). "Tensor networks for probabilistic sequence modeling". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 3079–3087.

Mitchell, J. and M. Lapata. (2010). "Composition in distributional models of semantics". *Cognitive science*. 34(8): 1388–1429.

Mitra, B. and N. Craswell. (2017). "Neural Text Embeddings for Information Retrieval".

Mitra, B. and N. Craswell. (2018). "An Introduction to Neural Information Retrieval". In: *An Introduction to Neural Information Retrieval*.

Mitra, M., C. Buckley, A. Singhal, and C. Cardie. (1997). "An Analysis of Statistical and Syntactic Phrases". In: *Computer-Assisted Information Searching on Internet. RIAO '97*. Montreal, Quebec, Canada. 200–214.

Miwa, M. (2018). "The Past, Present and Future of Information Retrieval: Toward a User-Centered Orientation".

Nakagawa, T., K. Inui, and S. Kurohashi. (2010). "Dependency tree-based sentiment classification using CRFs with hidden variables". In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 786–794.

Nielsen, M. A. and I. L. Chuang. (2011). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press.

Orus, R. (2014). "A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States". *Annals of Physics*. 349: 117–158.

Oseledets, I. V. (2011). "Tensor-train decomposition". *SIAM Journal on Scientific Computing*. 33(5): 2295–2317.

Pan, D., P. Zhang, J. Li, D. Song, J.-R. Wen, Y. Hou, B. Hu, Y. Jia, and A. De Roeck. (2013). "Using dempster-shafer's evidence theory for query expansion based on freebase knowledge". In: *Asia Information Retrieval Symposium*. Springer. 121–132.

Pan, W., W. Dong, M. Cebrian, T. Kim, J. H. Fowler, and A. S. Pentland. (2012). "Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems". *IEEE Signal Processing Magazine*. 29(2): 77–86.

Panahi, A., S. Saeedi, and T. Arodz. (2019). "word2ket: Space-efficient word embeddings inspired by quantum entanglement". *arXiv preprint arXiv:1911. 04975*.

Pang, B. and L. Lee. (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". *arXiv preprint cs/0506075*.

Pang, L., Y. Lan, J. Guo, J. Xu, and X. Cheng. (2016). "A Study of MatchPyramid Models on Ad-hoc Retrieval". *arxiv: org/abs/1606. 04648*.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. (2002). "Bleu: a Method for Automatic Evaluation of Machine Translation". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. 311–318. DOI: 10.3115/1073083.1073135.

Peng, Z., D. Song, Y. Hou, J. Wang, and P. Bruza. (2010). "Automata modeling for cognitive interference in users' relevance judgment". *Proc of Qi*: 125–133.

Pennington, J., R. Socher, and C. Manning. (2014). "Glove: Global Vectors for Word Representation". In: *Conference on Empirical Methods in Natural Language Processing*.

Peters, M. E., M. Neumann, L. Zettlemoyer, and W.-t. Yih. (2018). "Dissecting contextual word embeddings: Architecture and representation". *arXiv preprint arXiv:1808.08949*.

Piwowarski, B., I. Frommholz, Y. Moshfeghi, M. Lalmas, and C. J. Van. (2010a). "Filtering documents with subspaces". In: *In ECIR*.

Piwowarski, B., M. Lalmas, I. Frommholz, and K. van Rijsbergen. (2010b). "Exploring a Multidimensional Representation of Documents and Queries". In: *Adaptivity, Personalization and Fusion of Heterogeneous Information. RIAO '10*. Paris, France: LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE. 57–60.

Platonov, A., E. Poleschuk, I. Bessmertny, and N. Gafurov. (2018). "Using quantum mechanical framework for language modeling and information retrieval". In: *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE. 1–4.

Poria, S., E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency. (2017). "Context-dependent sentiment analysis in user-generated videos". In: *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*. 873–883.

Poria, S., D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea. (2018). "Meld: A multimodal multi-party dataset for emotion recognition in conversations". *arXiv preprint arXiv:1810.02508*.

Preller, A. (2007). "Towards discourse representation via pregroup grammars". In: *Journal of Logic Language Information*.

Revelli, A., M. M. Sala, G. Monaco, P. Becker, L. Bohatỳ, M. Hermanns, T. Koethe, T. Fröhlich, P. Warzanowski, T. Lorenz, *et al.* (2019). "Resonant inelastic x-ray incarnation of Young's double-slit experiment". *Science advances*. 5(1): eaav4020.

Richard, P. (1963). *The Feynman lectures on physics*. Addison-Wesley.

Robertson, S. E. (1977). "The probability ranking principle in IR". *Journal of documentation*. 33(4): 294–304.

Robertson, S. E. and S. Walker. (1994). "Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval". In: *SIGIR'94*. Springer. 232–241.

Ruthven, I. (2010). "Interactive information retrieval". *Annual Review of Information Science & Technology*. 42(1): 43–91.

Sadrzadeh, M. (2007). "High Level Quantum Structures in Linguistics and Multi Agent Systems." In: *AAAI Spring Symposium: Quantum Interaction*. 9–16.

Salton, G., C. S. Yang, and C. T. Yu. (2010). "A Theory of Term Importance in Automatic Text Analysis". *Journal of the American Society for Information ence & Technology*. 26.

Saracevic, T. (1996). "Relevance reconsidered". In: *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*. 201–218.

Schollwoeck, U. (2011). "The density-matrix renormalization group in the age of matrix product states". *Annals of Physics*. 326: 96–192.

Schuster, M. and K. K. Paliwal. (1997). "Bidirectional recurrent neural networks". *IEEE transactions on Signal Processing*. 45(11): 2673–2681.

Sennrich, R. and B. Haddow. (2016). "Linguistic input features improve neural machine translation". *arXiv preprint arXiv:1606.02892*.

Sennrich, R., B. Haddow, and A. Birch. (2016). "Edinburgh neural machine translation systems for wmt 16". *arXiv preprint arXiv:1606.02891*.

Severyn, A. and A. Moschitti. (2015). "Learning to rank short text pairs with convolutional deep neural networks". In: *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 373–382.

Shahroudy, A., W. Gang, and T. T. Ng. (2014). "Multi-modal feature fusion for action recognition in RGB-D sequences". In: *International Symposium on Communications*.

Shi, Y.-Y., L.-M. Duan, and G. Vidal. (2006). "Classical simulation of quantum many-body systems with a tree tensor network". *Physical review a*. 74(2): 022320.

Silverman, M. P. (2008). *Quantum superposition: counterintuitive consequences of coherence, entanglement, and interference*. Springer Science & Business Media.

Socher, R., A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. (2013). "Recursive deep models for semantic compositionality over a sentiment treebank". In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1631–1642.

Song, D., M. Lalmas, C. Rijsbergen, I. Frommholz, B. Piwowarski, J. Wang, P. Zhang, G. Zuccon, P. Bruza, S. Arafat, L. Azzopardi, E. Di Buccio, A. Huertas-Rosero, Y. Hou, M. Melucci, and S. Rueger. (2010). "How quantum theory is developing the field of Information Retrieval". Jan.

Song, F. and W. B. Croft. (1999). "A General Language Model for Information Retrieval". In: *Proceedings of the Eighth International Conference on Information and Knowledge Management. CIKM '99*. Kansas City, Missouri, USA: Association for Computing Machinery. 316–321. DOI: 10.1145/319950.320022.

Song, Q., W. Wang, W. Fu, Y. Sun, D. Wang, and Z. Gao. (2022). "Research on quantum cognition in autonomous driving". *Scientific reports*. 12(1): 1–16.

Sordoni, A., J. Y. Nie, and Y. Bengio. (2013). "Modeling term dependencies with quantum language models for IR". In: *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 653–662.

Srikanth, M. and R. Srihari. (2002). "Biterm Language Models for Document Retrieval". In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '02*. Tampere, Finland: Association for Computing Machinery. 425–426. DOI: 10.1145/564376.564476.

Srivastava, N. and R. Salakhutdinov. (2012). "Multimodal Learning with Deep Boltzmann Machines." In: *NIPS*. Vol. 1.

Stoudenmire, E. and D. J. Schwab. (2016). "Supervised learning with tensor networks". In: *Advances in Neural Information Processing Systems*. 4799–4807.

Surov, I. A., E. Semenenko, A. Platonov, I. Bessmertny, F. Galofaro, Z. Toffano, A. Y. Khrennikov, and A. Alodjants. (2021). "Quantum semantics of text perception". *Scientific Reports*. 11(1): 1–13.

Sutcliffe, A. and M. Ennis. (1998). "Towards a cognitive theory of information retrieval". *Interacting with computers*. 10(3): 321–351.

Tan, M., C. d. Santos, B. Xiang, and B. Zhou. (2015). "Lstm-based deep learning models for non-factoid answer selection". *arXiv preprint arXiv:1511.04108*.

Thierry, P. (2006). "Quantum computation and quantum information". *Mathematical Structures in Computer Science*. 17(6): 1115–1115.

Thouless, D. J. (2014). *The quantum mechanics of many-body systems*. Courier Corporation.

Uprety, S., D. Gkoumas, and D. Song. (2020). "A Survey of Quantum Theory Inspired Approaches to Information Retrieval". *ACM Comput. Surv.* 53(5). DOI: 10.1145/3402179.

Van Rijsbergen, C. J. (1979). *Information Retrieval*. Butterworth-Heinemann.

Van Rijsbergen, C. J. (2004). *The geometry of information retrieval*. Cambridge University Press.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is all you need". In: *Advances in neural information processing systems*. 5998–6008.

Verstraete, F., V. Murg, and J. Cirac. (2008). "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems". *Advances in Physics*. 57: 143–224.

Verstraete, F., M. M. Wolf, D. Pérez-García, and J. Cirac. (2006). "Criticality, the area law, and the computational power of projected entangled pair states." *Physical review letters*. 96 22: 220601.

Wan, S., Y. Lan, J. Guo, J. Xu, L. Pang, and X. Cheng. (2015). "A deep architecture for semantic matching with multiple positional sentence representations". *arXiv preprint arXiv:1511.08277*.

Wang, B., Q. Li, M. Melucci, and D. Song. (2019a). "Semantic Hilbert space for text representation learning". In: *The World Wide Web Conference*. 3293–3299.

Wang, B., P. Zhang, J. Li, D. Song, Y. Hou, and Z. Shang. (2016). "Exploration of quantum interference in document relevance judgement discrepancy". *Entropy*. 18(4): 144.

Wang, B., D. Zhao, C. Lioma, Q. Li, P. Zhang, and J. G. Simonsen. (2019b). "Encoding word order in complex embeddings". *arXiv preprint arXiv:1912. 12333*.

Wang, J., D. Song, P. Zhang, Y. Hou, and P. Bruza. (2010). "Explanation of relevance judgement discrepancy with quantum interference". In: *2010 AAAI Fall Symposium Series*.

Wang, L. (2016). "Discovering phase transitions with unsupervised learning". *Physical Review B*. 94: 195105.

Wang, L. (2017). "Exploring cluster Monte Carlo updates with Boltzmann machines". *Physical Review E*. 96(5). DOI: 10.1103/physreve.96.051301.

Wang, M., N. A. Smith, and T. Mitamura. (2007). "What is the Jeopardy model? A quasi-synchronous grammar for QA". In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. 22–32.

Wang, N., G. Gan, P. Zhang, S. Zhang, J. Wei, Q. Liu, and X. Jiang. (2022). "ClusterFormer: Neural Clustering Attention for Efficient and Effective Transformer". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics. 2390–2402. DOI: 10.18653/v1/2022.acl-long.170.

Wang, P., Z. Li, X. Pan, D. Ding, X. Chen, and Y. Hou. (2019c). "Density Matrix Based Preference Evolution Networks for E-Commerce Recommendation". In: *Database Systems for Advanced Applications - 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22-25, 2019, Proceedings, Part II*. Vol. 11447. *Lecture Notes in Computer Science*. Springer. 366–383. DOI: 10.1007/978-3-030-18579-4_22.

Wang, P., Z. Li, Y. Zhang, Y. Hou, and L. Ge. (2019d). "QPIN: A Quantum-inspired Preference Interactive Network for E-commerce Recommendation". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*. ACM. 2329–2332. DOI: 10.1145/3357384.3358076.

Wang, S. I. and C. D. Manning. (2012). "Baselines and bigrams: Simple, good sentiment and topic classification". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 90–94.

Wei, B., M. Wang, H. Zhou, J. Lin, J. Xie, and X. Sun. (2019). "Imitation learning for non-autoregressive neural machine translation". *arXiv preprint arXiv:1906.02041*.

White. (1992). "Density matrix formulation for quantum renormalization groups." *Physical review letters*. 69 19: 2863–2866.

Wiebe, J., T. Wilson, and C. Cardie. (2005). "Annotating expressions of opinions and emotions in language". *Language resources and evaluation*. 39(2): 165–210.

Williams, P., R. Sennrich, M. Nadejde, M. Huck, B. Haddow, and O. Bojar. (2016). "Edinburgh's statistical machine translation systems for wmt16". In: *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*. 399–410.

Xie, M., Y. Hou, P. Zhang, J. Li, W. Li, and D. Song. (2015). "Modeling Quantum Entanglements in Quantum Language Models". In: *Proceedings of the 24th International Conference on Artificial Intelligence. IJCAI'15*. Buenos Aires, Argentina: AAAI Press. 1362–1368.

Yang, C.-D. (2005). "Wave-particle duality in complex space". *Annals of Physics*. 319(2): 444–470.

Yang, H., M. Sloan, and J. Wang. (2015a). "Dynamic Information Retrieval Modeling". *Synthesis Lectures on Information Concepts Retrieval & Services*. 8(3): 409–410.

Yang, L., Q. Ai, J. Guo, and W. B. Croft. (2016). "aNMM: Ranking short answer texts with attention-based neural matching model". In: *Proceedings of the 25th ACM international on conference on information and knowledge management*. 287–296.

Yang, Y., W.-t. Yih, and C. Meek. (2015b). "Wikiqa: A challenge dataset for open-domain question answering". In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2013–2018.

Yang, Z., Z. Dai, R. Salakhutdinov, and W. W. Cohen. (2017). "Breaking the softmax bottleneck: A high-rank RNN language model". *arXiv preprint arXiv:1711.03953*.

Yu, L., K. M. Hermann, P. Blunsom, and S. Pulman. (2014). "Deep learning for answer sentence selection". *arXiv preprint arXiv:1412.1632*.

Zadeh, A., P. P. Liang, S. Poria, P. Vij, E. Cambria, and L.-P. Morency. (2018). "Multi-attention recurrent network for human communication comprehension". In: *Thirty-Second AAAI Conference on Artificial Intelligence*.

Zeng, W. and B. Coecke. (2016). "Quantum algorithms for compositional natural language processing". *arXiv preprint arXiv:1608.01406*.

Zhai, C. and J. Lafferty. (2017). "A study of smoothing methods for language models applied to ad hoc information retrieval". In: *ACM SIGIR Forum*. Vol. 51. No. 2. ACM New York, NY, USA. 268–276.

Zhang, J., P. Zhang, B. Kong, J. Wei, and X. Jiang. (2021a). "Continuous self-attention models with neural ODE networks". In: *Proc. of AAAI Conference on Aritificial Intelligence*.

Zhang, L., P. Zhang, X. Ma, S. Gu, Z. Su, and D. Song. (2019a). "A generalized language model in tensor space". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 7450–7458.

Zhang, P., H. Gao, Z. Hu, M. Yang, D. Song, J. Wang, Y. Hou, and B. Hu. (2022a). "A bias–variance evaluation framework for information retrieval systems". *Information Processing & Management*. 59(1): 102747.

Zhang, P., W. Hui, B. Wang, D. Zhao, D. Song, C. Lioma, and J. G. Simonsen. (2022b). "Complex-Valued Neural Network-Based Quantum Language Models". *ACM Trans. Inf. Syst.* 40(4). DOI: 10.1145/3505138.

Zhang, P., J. Niu, Z. Su, B. Wang, L. Ma, and D. Song. (2018a). "End-to-End Quantum-like Language Models with Application to Question Answering". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI'18*. New Orleans, Louisiana, USA: AAAI Press.

Zhang, P., Z. Su, L. Zhang, B. Wang, and D. Song. (2018b). "A Quantum Many-Body Wave Function Inspired Language Modeling Approach". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management. CIKM '18*. Torino, Italy: Association for Computing Machinery. 1303–1312. DOI: 10.1145/3269206.3271723.

Zhang, P., J. Zhang, X. Ma, S. Rao, G. Tian, and J. Wang. (2021b). "Text{TN}: Probabilistic Encoding of Language on Tensor Network". URL: https://openreview.net/forum?id=uUTx2LOBMV.

Zhang, S., P. Zhang, X. Ma, J. Wei, N. Wang, and Q. Liu. (2020a). "TensorCoder: Dimension-Wise Attention via Tensor Representation for Natural Language Modeling". *arXiv preprint arXiv:2008.01547*.

Zhang, Y., Q. Li, D. Song, P. Zhang, and P. Wang. (2019b). "Quantum-Inspired Interactive Networks for Conversational Sentiment Analysis". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization. 5436–5442.

Zhang, Y., D. Song, X. Li, and P. Zhang. (2018c). "Unsupervised sentiment analysis of twitter posts using density matrix representation". In: *European Conference on Information Retrieval*. Springer. 316–329.

Zhang, Y., D. Song, X. Li, P. Zhang, P. Wang, L. Rong, G. Yu, and B. Wang. (2020b). "A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis". *Information Fusion*. 62: 14–31.

Zhang, Y., D. Song, P. Zhang, P. Wang, J. Li, X. Li, and B. Wang. (2018d). "A quantum-inspired multimodal sentiment analysis framework". *Theoretical Computer Science*. 752: 21–40.

Zhao, X., P. Zhang, D. Song, and Y. Hou. (2011). "A Novel Re-ranking Approach Inspired by Quantum Measurement". In: *ECIR*.

Zi, C. (2017). "Application of machine learning methods in quantum many-body physics". *physics*.

Zuccon, G., L. Azzopardi, and C. Rijsbergen. (2009a). "The Quantum Probability Ranking Principle for Information Retrieval". In: *ICTIR*.

Zuccon, G., L. A. Azzopardi, and K. Van Rijsbergen. (2009b). "The quantum probability ranking principle for information retrieval". In: *Conference on the Theory of Information Retrieval*. Springer. 232–240.