

Natural Language Processing Meets Quantum Physics: A Survey and Categorization

Sixuan Wu^{1,2}, Jian Li³, Peng Zhang⁴ and Yue Zhang^{2,5}

¹Fudan University

²School of Engineering, Westlake University

³School of Science, Westlake University

⁴College of Intelligence and Computing, Tianjin University

⁵Institute of Advanced Technology, Westlake Institute for Advanced Study

Abstract

Recent research has investigated quantum NLP, designing algorithms that process natural language in quantum computers, and also quantum-inspired algorithms that improve NLP performance on classical computers. In this survey, we review representative methods at the intersection of NLP and quantum physics in the past ten years, categorizing them according to the use of quantum theory, the linguistic targets that are modeled, and the downstream application. The literature review ends with a discussion on the key factors to the success that has been achieved by existing work, as well as challenges ahead, with the goal of better understanding the promises and further directions.

1 Introduction

Quantum computing has received much interest in recent years. The basic idea is to make use the power of quantum mechanics for solving computational problems (Shor, 1999; Nielsen and Chuang, 2002). While particular quantum algorithms can be substantially faster alternatives to classical counterparts (Biamonte et al., 2017; Arute et al., 2019), the mathematical framework of quantum physics has also been exploited for cognition (Busemeyer and Bruza, 2012), optimization (Soleimanpour et al., 2014) and other disciplines. In the field of natural language processing (NLP), quantum mechanics has seen a surge of recent research interests, addressing problems ranging from lexical semantic ambiguities (Meyer and Lewis, 2020) to semantic composition (Coecke et al., 2020), and from information retrieval (Jiang et al., 2020) to text classification (Zhang et al., 2021), where different characteristics of quantum physics have inspired novel algorithms.

Despite its growing research literature, no survey has reviewed and categorized the quantum NLP field. The most relevant surveys are on quantum-inspired information retrieval (Uprety et al., 2020;

Melucci, 2015). However, they did not include many important findings in the quantum NLP field. Abohashima et al. (2020) and Garg and Ramakrishnan (2020) generally reviewed the field of quantum machine learning. They also briefly mentioned several quantum algorithms for NLP, but they did not discuss them comprehensively or in detail. The goal of our paper is to, for the first time, propose a categorization of quantum NLP in the past ten years, aiming to provide the latest knowledge of developments and achievements in this field.

We categorize existing work on quantum NLP based on the following three dimensions:

- 1 The types of algorithms. Many quantum-inspired NLP algorithms run on classical computers, and some quantum NLP algorithms can potentially be implemented on quantum hardware (Section 3).
- 2 The modeling target. Quantum physics is used for modeling different features of language (Section 4).
- 3 The applications. These algorithms have different applications, e.g. information retrieval, question answering (Section 5).

Although quantum NLP is still an emerging field, existing work shows exciting promise—not only better performance but also more efficient calculations are possible. In addition, noisy intermediate-scale quantum (NISQ) computers already exist and seem to have potential use in NLP tasks (Coecke et al., 2020; Lorenz et al., 2021). It has been shown that quantum NLP can take effect in addressing the inherent ambiguities of words, representing lexical semantic correlations, and calculating semantic composition, which is useful for a set of language modeling and information retrieval tasks. On the other hand, success has been achieved only on small scales, and the key reason for achieving

competitive performance still needs further understanding. The theoretical evidence provided in the literature cannot yet lead to a conclusion that quantum physics can gain substantial computational advantages in wider NLP tasks.

2 Quantum Physics Preliminaries

The simplest quantum mechanical system is a qubit, which has two possible states: $|0\rangle$ and $|1\rangle$, where $|\cdot\rangle$ is called the Dirac notation, and a ket $|\psi\rangle$ denotes a unit column vector. Similarly, the row vector ψ^\dagger is expressed as a bra $\langle\psi|$, where the dagger (\dagger) corresponds to the conjugate transpose. A qubit can be represented by the linear combination of states, often called **superposition**:

$$|\psi\rangle = a|0\rangle + b|1\rangle \quad (1)$$

where a and b are complex numbers and $|a|^2 + |b|^2 = 1$. Thus the state of a qubit is a unit vector in a two-dimensional complex vector space. When we measure a qubit we obtain either 0, with probability $|a|^2$, or 1, with probability $|b|^2$.

The superposition state can be used for representing multiple meanings of a word. For example, think of a mouse again as a small rodent and a hand-held pointing device. This two independent latent concepts can be denoted as $|\text{rodent}\rangle$ and $|\text{device}\rangle$. Then, the word 'mouse' can be modeled as a superposition state, i.e. $|\text{mouse}\rangle = a|\text{rodent}\rangle + b|\text{device}\rangle$.

Entanglement is another elementary and unique resource of quantum mechanics which plays a key role in many interesting applications of quantum computing. Consider the following two-qubit entangled Bell state (Nielsen and Chuang, 2002):

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} \quad (2)$$

As discussed earlier, when we measure the first qubit, we obtain two possible results: 0 with probability 1/2 and 1 with probability 1/2. According to Eq.2, a measurement of the second qubit always gives the same outcome as the measurement of the first qubit, because the measurement results of these two entangled qubits are correlated. Coecke et al. (2020) proposed that if words are encoded into quantum states, then the grammatical structure is to entangle these states. Because grammar is what correlates meanings between words. We will explain this in Section 4.2.2.

Projective measurements are the most general form of measurement in quantum mechanics, where the measurement operators are projectors P that satisfy $P^2 = P$. If the state is $|\psi\rangle$ before projective measurement then the probability that result m occurs is given by $p(m) = \langle\psi|P_m|\psi\rangle$. The state after measurement is:

$$|\psi\rangle \rightarrow \frac{P_m|\psi\rangle}{\sqrt{\langle\psi|P_m|\psi\rangle}} \quad (3)$$

Projective measurement can be applied to calculate cosine similarity in NLP, which measures the similarities between two vectors. Suppose $|A\rangle$ and $|B\rangle$ represent word A and B , respectively. Then, the cosine similarity of these two word vectors is

$$\cos^2(A, B) = \langle A|P_B|A\rangle = |\langle A|B\rangle|^2 \quad (4)$$

where $P_B = |B\rangle\langle B|$ is a projective measurement operator.

In addition to state vectors, quantum mechanics can also be formulated using **density matrix**, which is mathematically equivalent. Suppose that a quantum system is in one of the states $|\psi_i\rangle$, where i is an index, with probability p_i . The definition of the density matrix is:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad (5)$$

More information about quantum computing can be found in (Nielsen and Chuang, 2002).

3 The Types of Algorithms

Algorithms at the intersection of NLP and quantum physics can be implemented either on quantum computers or classical computers. The former ones are usually called quantum algorithms and the latter ones are usually named quantum-inspired or quantum-like models which are classical algorithms. We refer to both design of classical NLP algorithms inspired by quantum physics and quantum algorithms to process NLP tasks as quantum NLP.

We organize the main surveyed work in Table 1. This section and the next two sections discuss the categorization with regard to the algorithm type, the modeling target, and the application, respectively.

3.1 Quantum Algorithms

In quantum computing, a quantum algorithm is an algorithm that runs on real quantum computers. With regard to representation, Coecke et al.

| Paper | Types of Algorithms | Modeling Target | Applications |
|--------------------------------|---------------------|---------------------|--------------------------|
| (Zeng and Coecke, 2016) | Quantum | Word + Composition | / |
| (Meichanetzidis et al., 2020a) | Quantum | Word + Composition | / |
| (Meichanetzidis et al., 2020b) | Quantum | Word + Composition | Question answering |
| (Coecke et al., 2020) | Quantum | Word + Composition | Question answering |
| (Lorenz et al., 2021) | Quantum | Word + Composition | Syntactic classification |
| (Sordoni et al., 2013) | Classical | Composition | Information retrieval |
| (Xie et al., 2015) | Classical | Composition | Information retrieval |
| (Basile and Tamburini, 2017) | Classical | Composition | Speech recognition |
| (Li et al., 2018) | Classical | Composition | Information retrieval |
| (Zhang et al., 2018b) | Classical | Composition | Question answering |
| (Zhang et al., 2018a) | Classical | Composition | Question answering |
| (Zhang et al., 2018c) | Classical | Composition | Text classification |
| (Li et al., 2019) | Classical | Word + Composition | Question answering |
| (Zhang et al., 2019) | Classical | Composition | Text classification |
| (Lewis, 2019) | Classical | Word representation | / |
| (Meyer and Lewis, 2020) | Classical | Word representation | / |
| (Jiang et al., 2020) | Classical | Composition | Information retrieval |
| (Zhang et al., 2020) | Classical | Composition | Text classification |
| (Zhang et al., 2021) | Classical+Quantum | Word + Composition | Text classification |

Table 1: Categorization of main surveyed papers according to three dimensions defined in the Introduction.

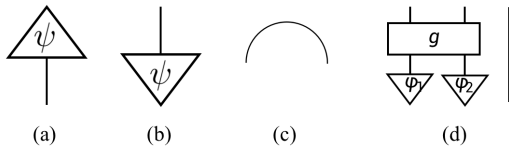


Figure 1: Graphical framework proposed by (Coecke and Kissinger, 2018; Coecke et al., 2020) to demonstrate: (a) A ket $|\psi\rangle$, (b) A bra $\langle\psi|$, (c) Bell state in Eq. 2, (d) $(g(|\varphi_1\rangle \otimes |\varphi_2\rangle)) \otimes I$, where matrix multiplication looks like connecting up the inputs and outputs of boxes and tensor product looks like placing boxes side by side.

(2010) constituted a graphical framework (DisCo-Cat) for natural language that combines words and builds the meaning of a sentence instead of thinking of a sentence as a bag of words. They devised a graphical framework from previous work which represents quantum mechanics pictorially by using lines, triangles, and so on (Coecke and Kissinger, 2018). As an example, in Figure 1, we use this graphical framework to demonstrate the ket, bra, and two-qubit entangled states introduced in Section 2.

Zeng and Coecke (2016) first discussed whether

a quantum computer can be applied to process natural language, showing a quantum algorithm for calculating sentence similarity that, under certain conditions, achieves a quadratic speedup over classical methods (see Table 2). This quadratic speedup, however, requires quantum random access memory (QRAM), which is expensive and remains unrealized (Biamonte et al., 2017). Considering this problem, Meichanetzidis et al. (2020a) and Coecke et al. (2020) proposed quantum algorithms that can potentially be implemented in existing NISQ computers. Wiebe et al. (2019) presented a representation for the linguistic structure which can encode NLP problems into small quantum devices. As a proof-of-concept experiment, Meichanetzidis et al. (2020b) performed the first quantum NLP task using a small dataset on NISQ hardware. To present larger-scale experiments, Lorenz et al. (2021) implemented models that solve sentence classification tasks on NISQ computers for datasets of size ≥ 100 sentences. These works pave the way for practical quantum NLP in the NISQ era.

| Algorithms | Typical cases |
|-----------------------|-----------------------------------|
| Classical Direct | $O(NM)$ |
| Classical Monte Carlo | $O(NMd^2r_{max}^4)$ |
| Quantum | $O(\sqrt{NM}\log(M)d^2r_{max}^4)$ |

Table 2: Complexity comparisons for different closest vector algorithms. Adapted from (Zeng and Coecke, 2016)

3.2 Classical Algorithms

Quantum-inspired or quantum-like NLP algorithms have been designed for classical computers, and some of them achieve comparable performance to state-of-the-art models (Jiang et al., 2020; Zhang et al., 2020). For the sake of applicability, these classical algorithms borrow mathematical frameworks from quantum mechanics but are not constrained by the quantum computing operations when processing the data.

Van Rijsbergen (2004) first proposed to unify information retrieval models into the mathematical framework of quantum mechanics in Hilbert space. Sordoni et al. (2013) proposed a quantum language model, which models term dependencies using the density matrix. This work indicates that the density matrix may be a more general representation of texts. Based on this, Basile and Tamburini (2017) presented a language model using the evolution of the state which can be implemented in speech recognition. Li et al. (2019) encoded words as quantum states and sentences as mixed systems.

Recently, in order to improve practicality, some quantum-inspired neural networks for natural language problems have been proposed. Zhang et al. (2019) use a density matrix based convolutional network to capture interactions within each utterance, outperforms a number of state-of-the-art sentiment analysis algorithms. Jiang et al. (2020) proposed a quantum interference inspired neural matching model with application to ad-hoc retrieval. The main difference between these quantum-inspired neural models for NLP and the existing neural based models is that the former models use the mathematical framework of quantum theory to describe language features. These features described by quantum theory are then used as the input of the neural network. Using quantum mechanics concepts to describe features have better interpretability, because they have more transparent physical explanations. It is also more beneficial to the subsequent neural network to extract useful information.

The above quantum-inspired neural networks are mainly for improving end-to-end performance, but still lack a theoretical foundation for the connection between quantum-inspired language model and neural network. Tensor networks, which factorize very large tensors into networks of smaller tensors, can help the theoretical understanding of existing neural networks (Levine et al., 2018). Based on tensor decomposition, Zhang et al. (2018b) proposed a quantum many-body wave function (QMWF) inspired language modeling and showed a mathematical understanding of using convolutional neural network (CNN). More recently, Zhang et al. (2021) proposed a tensor network method (namely TextTN) for natural language representation. Tensor network can not only run on a classical computer but also can be transformed into a quantum circuit. In addition, the hyper-parameters of TextTN can be well interpreted by the entanglement entropy (Zhang et al., 2021).

4 The Modeling Target

Both quantum-inspired algorithms and quantum algorithms can model different features in the language. We divide them into word representation (Section 4.1) and composition (Section 4.2).

4.1 Word Representation

How to represent words is essential for most NLP tasks and can affect performance. Using quantum physics for word representation has the potential to help including more features for words.

4.1.1 Modeling Word Ambiguity

Word ambiguity is a combination of distinct known meanings. Li et al. (2019), Li et al. (2018) and Coecke et al. (2020) adopted superposition state and complex number to formulate this combination. The latent concepts of a word form a set of pure orthonormal states of the space $\{|C_i\rangle\}$. This word t is modeled as a superposition state

$$|t\rangle = \sum_{i=1}^n a_i |C_i\rangle \quad (6)$$

in which the amplitude $\{a_i\}_{i=1}^n$ are complex numbers and $\sum_{i=1}^n |a_i|^2 = 1$. As mentioned in Section 2, if the superposition state is measured, it will collapse into the basis vector. This means that when a word is observed within a certain context it will collapse to one of its known meanings.

Li et al. (2019) and Wang et al. (2019) showed that we can benefit from the complex-valued word embedding and the phrases can be linked to some important features such as word positions. Moreover, the computational space increase exponentially with the size of the system (Coecke et al., 2020). If we consider a system of n qubits, then a quantum state of this system can represent a word that has 2^n latent concepts and is specified by 2^n amplitude. Trying to store all these complex numbers and vectors can be challenging on classical computers.

Meyer and Lewis (2020), Bankova et al. (2018) and Piedeleu et al. (2015) adopted density matrices to model lexical ambiguity. Unlike commonly-used methods which map words into vectors, they map words into matrices.

4.1.2 Modeling Hyponymy Relations

Hyponym refers to the fact that a word’s semantic field is included within another word’s. This relations can be encoded in projectors (Lewis, 2019; Bankova et al., 2018). For example, apple is an example of fruit, which is an example of food. This hyponymy relation can be encoded in projectors:

$$\begin{aligned} P_{apple} &= |00\rangle\langle 00| \\ P_{fruit} &= |00\rangle\langle 00| + |01\rangle\langle 01| \\ P_{food} &= |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10| \end{aligned} \quad (7)$$

here the normalized factors are ignored.

4.2 Composition

Given the meaning of each word, sentences can be understood by the composition of such lexical semantic units. Algorithms based on quantum physics can help to model this process.

4.2.1 Modeling Term Dependencies

Quantum-inspired algorithms have been considered for dependencies between terms in frequently occurring multiword expressions. The quantum language model (QLM) proposed by Sordoni et al. (2013) first applies quantum theory to model term dependencies, argue that there may be a situation in which classical probability fails and need to switch to a more general probabilistic theory. They map words w to projectors:

$$w \rightarrow \Pi_w = |e_w\rangle\langle e_w| \quad (8)$$

where $w \in V$ and $|e_w\rangle$ is the one-hot encoding of the word w . For example, consider $V =$

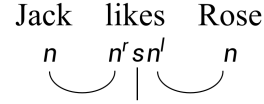


Figure 2: Diagrammatic form of the reduction $n(n^r s n^l)n \rightarrow (nn^r)s(n^l n) \rightarrow 1s1 \rightarrow s$, where n is noun, s is declarative statement, and cup denote the grammar reductions. According to pregroup grammar (Lorenz et al., 2021; Lambek, 2008), Jack likes Rose is grammatical because of above reduction.

$\{natural, language\}$. Then $\Pi_{language}$ is:

$$\Pi_{language} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (9)$$

The relationship linking two or more words is represented by a subset of the vocabulary $\kappa = \{w_1, w_2, \dots, w_n\}$ and encoded into a new projector:

$$\kappa \rightarrow |\kappa\rangle\langle \kappa|, |\kappa\rangle = \sum_{i=1}^n a_i |e_{w_i}\rangle \quad (10)$$

where $\{a_i\}_{i=1}^n$ are real numbers and $\sum_{i=1}^n a_i^2 = 1$. For example, we can model the dependency between natural and language, $\kappa_{nl} = \{natural, language\}$, by $K_{nl} = |\kappa_{nl}\rangle\langle \kappa_{nl}|$, where $|\kappa_{nl}\rangle = \sqrt{\frac{2}{5}}|e_{natural}\rangle + \sqrt{\frac{3}{5}}|e_{language}\rangle$. Then

$$K_{nl} = \begin{bmatrix} \frac{2}{5} & \frac{\sqrt{6}}{5} \\ \frac{\sqrt{6}}{5} & \frac{3}{5} \end{bmatrix} \quad (11)$$

$|\kappa_{nl}\rangle$ is a superposition state and K_{nl} is a density matrix. In quantum mechanics, elements of the density matrix K_{nl} contain the correlation between quantum states $|e_{natural}\rangle$ and $|e_{language}\rangle$, thus dependency between natural and language is modeled. This method of modeling term dependency is interpretable and has physical meaning.

Some algorithms have been proposed based on above QLM. Xie et al. (2015) took entanglement into consideration which is not considered in original QLM, Zhang et al. (2020) adopted word embedding instead of one-hot encoding, and so on. The basic and important idea behind these algorithms is to treat word vectors as quantum states from which we can obtain the density matrix of the sentence or document. Then this density matrix naturally contains the correlation of these quantum states, which means the dependence between words is modeled.

| Task | Dataset | Models | Metrics | |
|----------------------------|---|---------------------------------------|----------|----------|
| Text classification | MELD dataset Sentiments (3-class) | CNN (Kim, 2014) | F1 | Accuracy |
| | | RoBERTa (Liu et al., 2019) | 0.604 | 0.609 |
| | | QIN (Zhang et al., 2019) | 0.721 | - |
| | | QMN (Zhang et al., 2020) | 0.662 | 0.679 |
| | | Doc2vector (Le and Mikolov, 2014) | 0.729 | 0.756 |
| | OMD dataset | SentiStrength (Thelwall et al., 2010) | 0.3979 | 0.6103 |
| | | GQLM+QRE (Zhang et al., 2018c) | 0.6352 | 0.6110 |
| | | | 0.6261 | 0.6298 |
| | SST-5 | | Accuracy | |
| | | Star-Transformer (Guo et al., 2019) | 53.0 | |
| BERT (Devlin et al., 2018) | | 52.9 | | |
| | BERT+TextTN (Zhang et al., 2021) | 54.8 | | |
| Question answering | WIKIQA | | MAP | MRR |
| | | Bigram-CNN (Yu et al., 2014) | 0.6190 | 0.6281 |
| | | AP-BILSTM (Santos et al., 2016) | 0.6705 | 0.6842 |
| | | NNQLM-II (Zhang et al., 2018a) | 0.6496 | 0.6594 |
| | | CNM (Li et al., 2019) | 0.6748 | 0.6864 |
| Information retrieval | TREC 2013 | | MAP@10 | NDCG@10 |
| | | Unigram | 4.91 | 6.05 |
| | | QLM (Sordoni et al., 2013) | 6.14 | 6.70 |
| | | QLM-QE (Li et al., 2018) | 8.94 | 10.37 |
| | ClueWeb-09-Cat-B | | MAP | NDCG@20 |
| | | MP (Pang et al., 2016) | 0.066 | 0.158 |
| | | Conv-KNRM (Dai et al., 2018) | 0.121 | 0.285 |
| | | QLM (Sordoni et al., 2013) | 0.082 | 0.164 |
| | QINM (Jiang et al., 2020) | 0.134 | 0.338 | |

Table 3: Tasks and datasets that quantum NLP has shown comparable or better performance over traditional NLP. **The quantum NLP algorithms** are indicated in **bold**.

4.2.2 Modeling Grammar

Pregroup gramma (Lambek, 1997) is used for analyzing the structure of natural languages. As an algebraic gadget, pregroup grammar can be denoted using having cup-shaped wires (Lambek, 2008). We show an example sentence in Figure 2. From Figure 1 and Figure 2, we can see that diagrammatic frameworks used for quantum mechanics and pregroup grammar are partially similar. Coecke et al. (2010) introduced a model based on tensor product composition, which uses pregroup grammar to compute the meaning of sentences and phrases. Coecke et al. (2020) recast this model in quantum computational terms and showed that pregroup can always be made using only Bell-effect and identities.

Here is an example of how to use Bell-effect and identities to represent applying an adjective to a noun. Assuming the meaning of story is a 1-qubit state $|\psi_{story}\rangle \in \mathbb{C}^2$ and the meaning of adjective happy is a 2-qubit state $|\psi_{love}\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$. In happy story, happy modifies the noun story. Coecke et al. (2020) model this modification using

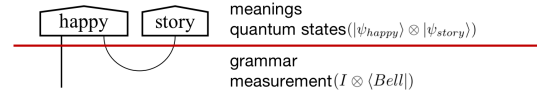


Figure 3: Diagrammatic notation of happy story

$|\psi_{happy}\rangle \otimes |\psi_{story}\rangle$, calculating:

$$|\psi_{happystory}\rangle = (I \otimes \langle Bell|) \cdot (|\psi_{happy}\rangle \otimes |\psi_{story}\rangle) \quad (12)$$

where $\langle Bell| = \langle 00| + \langle 11|$ and I is the identity. The mapping $(I \otimes \langle Bell|)$ shows the interaction between the meaning of words. Using diagrammatic notation (Coecke and Kissinger, 2018; Coecke et al., 2020), our example is illustrated in Figure 3. The pentagon represents the quantum state, the straight line represents the identity matrix, and the cup-shaped wire represents the Bell-effect. Coecke et al. (2020) also showed that this type of wire structure and pregroup grammar can be equivalent, and thus to some extent NLP is quantum native.

5 Applications

Quantum NLP shows comparable or better performance compared with strong baselines for some

tasks. We summarize the results of these algorithms in Table 3.

Information retrieval (IR). Sordoni et al. (2013) first proposed a quantum language model for IR, representing terms in queries and documents as superposition events attached with quantum probability, which has no classical analog. Extensions of the quantum language model have also been proposed for IR. Xie et al. (2015) advanced the QLM framework by taking into account quantum entanglement, which has a significant cognitive implication. Li et al. (2018) proposed an algorithm to help improve convergence. Jiang et al. (2020) took interference into account, which produces additional contributions to the total probability beyond classical cases. Based on this new contributions, they proposed a matching model for ad-hoc retrieval. The quantum matching models outperform some traditional models.

Question answering (QA). Zhang et al. (2018a) used density matrices to represent questions and answers and introduced a joint representation to model the similarities between the question and answers. This joint representation is then used as an input to a neural network. Li et al. (2019) proposed a complex-valued network for QA, which is interpretable and shows comparable performance to strong CNN and RNN baselines. Coecke et al. (2020) mentioned that QA tasks can be executed on quantum computers. After mapping a question to a vector, QA tasks become the task to find the closest vector in the answer vectors pool. They exploited quantum advantage for finding the closest vector (Wiebe et al., 2015) and showed quantum speedup. Meichanetzidis et al. (2020b) showed the first-ever quantum NLP experiment on quantum hardware through a QA task. Although this is a proof-of-concept experiment, it paved the way for the future use of quantum computers to deal with practical NLP problems.

Speech recognition. Basile and Tamburini (2017) introduced a quantum language model with the application for speech recognition, where words are encoded into measurement operators and the sequence of words is modeled as the evolution of quantum systems.

Text classification. Zhang et al. (2018c) explored the possibility of using quantum physics on sentiment classification tasks. Two sentiment dic-

tionaries were constructed. They generated density matrices for dictionaries and documents and used quantum relative entropy as characterization of the similarity between dictionaries and documents to determine its sentiment. Zhang et al. (2019) introduced quantum-inspired interactive networks, where a density matrix that capture correlations between words was used as an input of long short-term memory neural network. In order to effectively combine multiple information from different sources, Zhang et al. (2020) further extended their work with two modalities, namely text and visual modalities. Considering the interpretability and expressive power of tensor network, Zhang et al. (2021) proposed a tensor network based architecture for natural language.

6 Discussion

6.1 Benefits

In this section, we make a summary of potential benefits of quantum NLP, and discuss the most salient directions that remain under-explored due to various reasons.

Lowering computational cost. Some articles have demonstrated quantum speedup for specific NLP tasks, such as question answering (Coecke et al., 2020; Zeng and Coecke, 2016). Quantum search algorithm (Grover, 1996), quantum nearest-neighbor algorithm (Wiebe et al., 2015) and other quantum algorithms which achieve speedup over classical algorithms could be used after classical language features are encoded into quantum states. As mentioned in Section 4, the quantum superposition is suitable for modeling uncertainties in language, such as word ambiguity (Li et al., 2019; Wang et al., 2019). And entanglement can describe the composition of lexical semantic units (Coecke et al., 2020; Meichanetzidis et al., 2020b). It's possible that, by adaptations to quantum algorithms and deployments to quantum computers, a family of NLP tasks can enjoy quantum speedup.

Enhancing learning ability. Quantum mechanics is well-known to generate counter-intuitive patterns (Biamonte et al., 2017). It is reasonable to hope that quantum computers can recognize some patterns that cannot be recognized by classical computers. As shown in Table 3, some quantum NLP models have shown comparable or better performance over strong baselines. And the framework of quantum mechanics can be applied to model some features that are difficult to model with classical probabil-

ity. For example, quantum theory is used to model interference phenomenon in information retrieval (Jiang et al., 2020) and term dependencies (Sordani et al., 2013). It's more consistent with human cognition. Li et al. (2021) demonstrate that neural machine translation models fail badly on compositional generalization. According to existing paper, we believe quantum NLP models have potential advantages in compositional generalization problem.

Increasing storage capacity. Quantum computers have strong storage capabilities. As mentioned before, Coecke hold the view that NLP is quantum-native (Meichanetzidis et al., 2020b; Coecke et al., 2020) such that the exponentially large vector space required to represent sentences can only be naturally and feasibly realized in quantum computers. From this point of view, developments of quantum language models will be beneficial also in terms of storage efficiency.

6.2 Future Directions

Despite the emerging promises, quantum NLP has yet to see its full-fledged advantages to the dominant neural methods. Significant advances in one or more of the following directions can give strong boosts to the research field.

Quantum machine learning. Existing work showed that there is a fundamental connection between machine learning and quantum physics (Levine et al., 2018; Hughes et al., 2019). For example, tensor network is a method that bridges machine learning and quantum theory, which can also enhance the theoretical understanding of existing neural networks (Levine et al., 2019). For NLP, designing an effective tensor network approach can lead to better interpretability (Zhang et al., 2021). On the other hand, most quantum NLP models still use real vectors (Jiang et al., 2020; Zhang et al., 2020), partly because there are no obvious features corresponding to the imaginary part. However, quantum phenomena cannot be fully expressed without complex numbers. In quantum neural networks, complex numbers and quantum phenomena can be naturally modeled. It has been shown that both complex-valued representation of natural language (Li et al., 2019; Wang et al., 2019), and complex-valued neural networks (Trabelsi et al., 2018) can lead to benefits.

Wider applications. We have shown that quantum NLP algorithms can be used for information retrieval (Jiang et al., 2020), question answering

(Meichanetzidis et al., 2020b), and so on. These are relatively simple tasks and quantum NLP models have not been extended to more challenging tasks such as text generation and automatic summarization. Finding wider NLP tasks that can benefit from quantum physics is also a remaining direction.

Quantum advantages. In quantum computing, quantum supremacy or quantum advantage is the goal of demonstrating that a quantum computer can solve a problem that no classical computer can solve in any reasonable amount of time. Whether there are concrete examples in NLP that can show quantum advantages is a fundamental and important question. According to existing paper, there may be quantum advantages in NLP tasks which need similarity calculations, such as the similarity of the query and the documents, the similarity of the sentences, and the similarity of the question and the answers.

7 Conclusion

Thus far, articles have demonstrated early success in representing and processing text using quantum computers. Their design is scalable and when hardware becomes more powerful they can scale up the size of the meaning spaces and complexity of the tasks. The key to whether or not quantum computers will be used to deal with NLP in the future lies in whether quantum algorithms can show quantum advantage. Meanwhile, quantum-inspired models have shown strong performance on classical computers for certain tasks, and have better interpretability. The main difficulty in this direction is that neural networks have already achieved high accuracy on many NLP tasks. Nevertheless, it is still worthwhile to explore the mathematical framework of quantum mechanics where a strong expressive ability and a corresponding physical explanation are expected. Finally, it can also be possible that if neither of the above two directions has major breakthroughs, then this quantum NLP field may temporarily lose research attention during a period of time.

Acknowledgements

Yue Zhang is the corresponding author. We thank the anonymous reviewers for their insightful comments and suggestions.

References

- Zainab Abohashima, Mohamed Elhosen, Essam H Houssein, and Waleed M Mohamed. 2020. Classification with quantum machine learning: A survey. *arXiv preprint arXiv:2006.12270*.
- Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. 2019. Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510.
- Dea Bankova, Bob Coecke, Martha Lewis, and Dan Marsden. 2018. Graded hyponymy for compositional distributional semantics. *Journal of Language Modelling*, 6(2):225–260.
- Ivano Basile and Fabio Tamburini. 2017. Towards quantum language models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1840–1849.
- Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature*, 549:195–202.
- Jerome R Busemeyer and Peter D Bruza. 2012. *Quantum models of cognition and decision*. Cambridge University Press.
- Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. 2020. Foundations for near-term quantum natural language processing. *arXiv preprint arXiv:2012.03755*.
- Bob Coecke and Aleks Kissinger. 2018. Picturing quantum processes. In *International Conference on Theory and Application of Diagrams*, pages 28–31. Springer.
- Bob Coecke, Mehrnoosh Sadzadeh, and Stephen J Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis*, 36:345–384.
- Zhuyun Dai, Chenyan Xiong, Jamie Callan, and Zhiyuan 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*, pages 126–134.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Siddhant Garg and Goutham Ramakrishnan. 2020. Advances in quantum deep learning: An overview. *arXiv preprint arXiv:2005.04316*.
- Lov K Grover. 1996. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219.
- Qipeng Guo, Xipeng Qiu, Pengfei Liu, Yunfan Shao, Xiangyang Xue, and Zheng Zhang. 2019. Star-transformer. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1315–1325.
- Tyler W Hughes, Ian AD Williamson, Momchil Minkov, and Shanhui Fan. 2019. Wave physics as an analog recurrent neural network. *Science advances*, 5(12):eaay6946.
- Yongyu Jiang, Peng Zhang, Hui Gao, and Dawei 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*, pages 19–28.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751. Association for Computational Linguistics.
- Joachim Lambek. 1997. Type grammar revisited. In *International conference on logical aspects of computational linguistics*, pages 1–27. Springer.
- Joachim Lambek. 2008. *From Word to Sentence: a computational algebraic approach to grammar*. Polimetrica sas.
- Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.
- Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. 2019. Quantum entanglement in deep learning architectures. *Physical review letters*, 122(6):065301.
- Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. 2018. Deep learning and quantum entanglement: Fundamental connections with implications to network design. In *International Conference on Learning Representations (ICLR)*.
- Martha Lewis. 2019. Compositional hyponymy with positive operators. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 638–647.
- Qiuchi Li, Massimo Melucci, and Prayag Tiwari. 2018. Quantum language model-based query expansion. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 183–186.
- Qiuchi Li, Benyou Wang, and Massimo Melucci. 2019. Cnm: An interpretable complex-valued network for matching. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 4139–4148.

- Yafu Li, Yongjing Yin, Yulong Chen, and Yue Zhang. 2021. On compositional generalization of neural machine translation. *arXiv preprint arXiv:2105.14802*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Robin Lorenz, Anna Pearson, Konstantinos Meichanetzidis, Dimitri Kartsaklis, and Bob Coecke. 2021. Qnlp in practice: Running compositional models of meaning on a quantum computer. *arXiv preprint arXiv:2102.12846*.
- Konstantinos Meichanetzidis, Stefano Gogioso, Giovanni De Felice, Nicolò Chiappori, Alexis Toumi, and Bob Coecke. 2020a. Quantum natural language processing on near-term quantum computers. *arXiv preprint arXiv:2005.04147*.
- Konstantinos Meichanetzidis, Alexis Toumi, Giovanni de Felice, and Bob Coecke. 2020b. Grammar-aware question-answering on quantum computers. *arXiv preprint arXiv:2012.03756*.
- Massimo Melucci. 2015. *Introduction to information retrieval and quantum mechanics*. Springer.
- Francois Meyer and Martha Lewis. 2020. Modelling lexical ambiguity with density matrices. *arXiv preprint arXiv:2010.05670*.
- Michael A Nielsen and Isaac Chuang. 2002. Quantum computation and quantum information.
- Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng. 2016. A study of matchpyramid models on ad-hoc retrieval. *arXiv preprint arXiv:1606.04648*.
- Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke, and Mehrnoosh Sadrzadeh. 2015. Open system categorical quantum semantics in natural language processing. In *6th Conference on Algebra and Coalgebra in Computer Science*, page 270.
- Cicero dos Santos, Ming Tan, Bing Xiang, and Bowen Zhou. 2016. Attentive pooling networks. *arXiv preprint arXiv:1602.03609*.
- Peter W Shor. 1999. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332.
- Mohadeseh Soleimanpour, Moghadam, Hossein Nezamabadi-Pour, and Malihe M Farsangi. 2014. A quantum inspired gravitational search algorithm for numerical function optimization. *Information Sciences*, 267:83–100.
- Alessandro Sordani, Jian-Yun Nie, and Yoshua Bengio. 2013. Modeling term dependencies with quantum language models for ir. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 653–662.
- Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558.
- Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. 2018. Deep complex networks. In *International Conference on Learning Representations (ICLR)*.
- Sagar Uprety, Dimitris Gkoumas, and Dawei Song. 2020. A survey of quantum theory inspired approaches to information retrieval. *ACM Computing Surveys (CSUR)*, 53(5):1–39.
- Cornelis Joost Van Rijsbergen. 2004. *The geometry of information retrieval*. Cambridge University Press.
- Benyou Wang, Donghao Zhao, Christina Lioma, Qiuchi Li, Peng Zhang, and Jakob Grue Simonsen. 2019. Encoding word order in complex embeddings. In *International Conference on Learning Representations (ICLR)*.
- Nathan Wiebe, Alex Bocharov, Paul Smolensky, Matthias Troyer, and Krysta M Svore. 2019. Quantum language processing. *arXiv preprint arXiv:1902.05162*.
- Nathan Wiebe, Ashish Kapoor, and Krysta M Svore. 2015. Quantum nearest-neighbor algorithms for machine learning. *Quantum information and computation*, 15.
- Mengjiao Xie, Yuexian Hou, Peng Zhang, Jingfei Li, Wenjie Li, and Dawei Song. 2015. Modeling quantum entanglements in quantum language models. In *International Joint Conferences on Artificial Intelligence (IJCAI)*.
- Lei Yu, Karl Moritz Hermann, Phil Blunsom, and Stephen Pulman. 2014. Deep learning for answer sentence selection. *arXiv preprint arXiv:1412.1632*.
- William Zeng and Bob Coecke. 2016. Quantum algorithms for compositional natural language processing. *arXiv preprint arXiv:1608.01406*.
- Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. 2018a. End-to-end quantum-like language models with application to question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Peng Zhang, Zhan Su, Lipeng Zhang, Benyou Wang, and Dawei 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*, pages 1303–1312.

- Peng Zhang, Jing Zhang, Xindian Ma, Siwei Rao, Guangjian Tian, and Jun Wang. 2021. Text{tn}: Probabilistic encoding of language on tensor network.
- Yazhou Zhang, Qiuchi Li, Dawei Song, Peng Zhang, and Panpan Wang. 2019. Quantum-inspired interactive networks for conversational sentiment analysis. In *International Joint Conferences on Artificial Intelligence (IJCAI)*.
- Yazhou Zhang, Dawei Song, Xiang Li, and Peng Zhang. 2018c. Unsupervised sentiment analysis of twitter posts using density matrix representation. In *European Conference on Information Retrieval*, pages 316–329. Springer.
- Yazhou Zhang, Dawei Song, Xiang Li, Peng Zhang, Panpan Wang, Lu Rong, Guangliang Yu, and Bo Wang. 2020. A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis. *Information Fusion*, 62:14–31.