Modeling Natural Language via Quantum Many-body Wave Function and Tensor Network

Yitong Yao
College of Intelligence and Computing
Tianjin University
Tianjin, China
yitongyao@tju.edu.cn

Peng Zhang*
College of Intelligence and Computing
Tianjin University
Tianjin, China
pzhang@tju.edu.cn

Jing Zhang
College of Intelligence and Computing
Tianjin University
Tianjin, China
zhang_jing@tju.edu.cn

Abstract

Quantum machine learning has attracted increasing attention and successful applications in processing natural images. Compared with image processing or computer vision, the natural language processing (NLP) area, would have more fundamental connections with quantum theory and models. In this paper, we will briefly introduce recent developments on the quantum-inspired language models. Particularly, we will describe two recent language models using quantum many-body wave function and tensor network, respectively, as the theoretical formalism.

1 Introduction

Recently, the profound idea and sound formulation of quantum theory have been applied in machine learning and formed an interdisciplinary field, known as quantum machine learning [3, 13]. Quantum machine learning has achieved promising developments in processing and classifying natural images [15, 10, 16]. Here, we argue that compared with images, the textual languages, could have more fundamental connections with the quantum physics, and have more challenging problems.

First, in terms of the uncertainty modeling, compared with pixels (as the basic units in an image), the words (as the basic units of language) can have multiple semantic meanings simultaneously, before we measure them (e.g., putting it in a context), so that the word state (even a sentence state) can be represented by a superposition state in quantum mechanics [4]. Second, in terms of the cognitive aspect, the image processing and recognition are more towards the perception level, but the language processing and understanding are more towards the cognitive level. Recently, research studies have discovered the natural correspondence between the cognitive science and quantum theory [5]. Therefore, it is essential to study the machine learning and artificial intelligence problems from a quantum cognition points of view.

Despite of the rationality of modeling languages using quantum mechanics, it is challenging to develop effective and practical quantum-inspired language models. Compared with the continuous value of a pixel representation, the discrete nature of a word representation has long been the obstacle of NLP problems. The recent developed distributed representations (e.g., word embeddings) have

*the corresponding author

First Workshop on Quantum Tensor Networks in Machine Learning, 34th Conference on Neural Information Processing Systems (NeurIPS 2020).

alleviated this problem to some extent, but the relatively high dimensionality of word vectors still can make it more difficult to represent the a textual sentence than an image, using quantum tools, e.g., tensor network. On the other hand, the semantic dependencies between words, which need more hidden information, are difficult to learn directly from spatial features (e.g., distance in images).

In natural language processing and information retrieval areas, researchers are devoted to build principled and effective quantum-inspired models [1]. A milestone is the work named as a quantum language model (QLM), proposed by Sordoni, Nie and Bengio [14], which aims to model the term dependency in a principled manner. To avoid the disadvantages of tradition language models (LMs), they attempted to use density matrix to encode the probability measurement for both single words and compound words [19]. QLM is for the first time effectively apply the quantum probability measurement theory in modeling the language , and performed effectively in the ad-hoc information retrieval task. After QLM, neural network are utilized to to further improve the practicality of the quantum language models [18]. Such a Neural Network based Quantum-like Language Model (NNQLM), developed a more effective approach to the density matrix, built QLM in an end to end neural network, and achieved promising empirical results on the question answering (QA) tasks.

However, the achievements of QLM and NNQLM were innovative but still limited. The models did not explore and apply the fundamental theoretical connections between quantum theory and neural network [19], in the language modeling process. Therefore, we would like to describe two recent language models using quantum many-body wave function [19] and tensor network [17], respectively.

2 A Language Modeling Approach based on Quantum Many-body System

Inspired by the fundamental connections between quantum theory and natural language processing, a new approach called Quantum Many-body Wave Function inspired Language Modeling (QMWF-LM)[19] is proposed. It regards different words as diverse particles in a many-body system, and a word containing multiple meanings equals to a particle lying in a superposition [4]. Thus it is reasonable to directly use the wave function to represent the words in order to express the complex interactions between words with different meanings.

For a word with multiple meanings, QMWF-LM sets each meaning as a basic vector $|\phi_{h_i}\rangle$, and the word x_i could use quantum many-body wave function to express as:

$$|x_i\rangle = \sum_{h_i=1}^{M} \alpha_{i,h_i} |\phi_{h_i}\rangle \tag{1}$$

where α_{i,h_i} could be regarded as the probability of word x_i with a meaning of $|\phi_{h_i}\rangle$. For a sentence $S = [x_1, x_2, \dots, x_N]$, using tensor product to multiply each word, it would represent a semantic space which contains every kind of combination of meanings and can be represented as

$$|\psi_S^{ps}\rangle = |x_1\rangle \otimes \ldots \otimes |x_N\rangle = \sum_{h_1,\ldots,h_N=1}^{M} \mathcal{A}_{h_1\ldots h_N} |\phi_{h_1}\rangle \otimes \ldots \otimes |\phi_{h_N}\rangle \tag{2}$$

To calculate the expression of the whole corpus, QMWF-LM multiplies all the sentences using tensor product, turning the expression into $|\psi_S\rangle$, and the coefficient tensor turns into $\mathcal{T}_{h_1\ldots h_N}$. Thus, the projection from the whole corpus to a specific sentence is expressed using inner product as $\langle\psi_S^{ps} \mid \psi_S\rangle$. Though all the combined meanings could be expressed, the drawback of the approach is that there are a lot of useless combinations, and the result is that the dimensionality of the projection is too high to compute. So it is necessary to execute a low-rank approximation to cut off the extra space and make the Eq.2 computable. Therefore, the approach uses Canonical Polyadic Decomposition (CP decomposition [7]) method, inputting $\boldsymbol{x}_i = (\alpha_{i,h_1}, \ldots, \alpha_{i,h_M})^T$, now the equation is as follows:

$$\langle\psi_S^{ps} \mid \psi_S\rangle = \sum_{h_1,\ldots,h_N=1}^{M} \mathcal{T}_{h_1\ldots h_N} \times \mathcal{A}_{h_1\ldots h_N} = \sum_{r=1}^{R} t_r \prod_{i=1}^{N} \left(\sum_{h_i=1}^{M} e_{r,i,h_i} \cdot \alpha_{i,h_i} \right) \tag{3}$$

where e_{r,i,h_i} is an element of $e_{r,i} = (e_{r,i,1}, \ldots, e_{r,i,M})^T$, which is a unit vector with M dimensions decomposed from global tensor $\mathcal{T}$. In addition, during the process of mathematical derivation, it is surprising to realize that the format of the final projection from local representation of a sentence to the

Table 1: Projection using CNN architecture

Input	$\boldsymbol{x}_i = (\alpha_{i,h_1}, \ldots, \alpha_{i,h_M})^T$
Convolution	$\Sigma_{r,i} = \sum_{h_i=1}^{M} e_{r,i,h_i} \cdot \alpha_{i,h_i}$
Product pooling	$\Pi_r = \prod_{i=1}^{N} \Sigma_{r,i}$
Output	$\sum_{r=1}^{R} t_r \cdot \Pi_r$

global representation of corpus could be perfectly matched with the architecture of a Convolutional Neural Network (CNN) with product pooling [6]. Table 1 shows the correspondence between CNN and QMWF-LM. Therefore, QMWF-LM could not only utilize the architecture of CNN to compute the complex projection, but also provide a deeper understanding from a new perspective.

To examine the effectiveness of the approach, QMWF-LM is successfully leveraged into text matching field, especially the Question Answering (QA) tasks. Moreover, the experimental performance proves the correctness of QMWF-LM, which also shows that it is feasible to use quantum many-body system to simulate natural language system.

It is a breakthrough to make analogy between quantum many-body system with language semantic space. QMWF-LM proved the feasibility of using quantum many-body wave function to derive a new word representation. Also, the approach can be used to explain the design of CNN, which improves the interpretability of neural network. Nevertheless, though it is accessible to encode the sentence features with the help of CNN and testify its effectiveness in QA tasks successfully, QMWF-LM cannot be regarded as a language model since a language model needs to work out the joint probability which the approach has not involved. Therefore, QMWF-LM is just a language modeling approach promoting a new representation of words and sentences using quantum many-body wave function.

3 Generalized Language Model using Tensor Network

Since QMWF-LM [19] has proved the usefulness of leveraging quantum wave function to construct the words and sentences in describing interactions among context, a more generalized language model called Tensor Space Language Model(TSLM)[17] is introduced, using Eq.1 to represent words and sentences and introduce the method of tensor network.

A goal of language modeling is to learn the joint probability function of sequences of words in a language[2]. Based on the word and sentence representation using quantum wave function, it is easy to find out the mathematical expression of joint probability and the model uses $\boldsymbol{c}$ to represent the assembly of all sentences, while s_i represents a specific sentence. As the expressions derived in the previous section, the probability of s_i appearing in mixed representation $\boldsymbol{c}$ is:

$$p(s_i) = \langle s_i, \boldsymbol{c} \rangle = \sum_{d_1,\ldots,d_n=1}^{m} \mathcal{T}_{d_1 \ldots d_n} \mathcal{A}_{d_1 \ldots d_n} \tag{4}$$

$\boldsymbol{c}$ can be regarded as the total sampling distribution, and the tensor inner product represents probability that a sentence s appears in a language. Furthermore, the conditional probability distribution is:

$$p\left(w_t \mid w_1^{t-1}\right) = \operatorname{softmax}\left(\left\langle \mathcal{T}_{(t)}, \mathcal{A}_{(t-1)} \right\rangle\right) \tag{5}$$

where w_1^{t-1} represents $(w_1, w_2, \ldots, w_{t-1})$. Since $\mathcal{T}$ is a high dimension tensor, low-rank approximation on it using tensor-train decomposition [12] is needed, shown in Figure 1(a). Besides, $\mathcal{A}$ is a essentially rank-one so it could be directly decomposed into production of a list of rank-one tensor $\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \otimes \cdots \otimes \boldsymbol{\alpha}_n$. Figure 1(b) shows the whole process of using tensor network to recursively calculate the conditional probability of TSLM, and the expressions are as follows:

$$p\left(w_t \mid w_1^{t-1}\right) = \operatorname{softmax}(\boldsymbol{y}_t) \tag{6}$$

$$y_t = V h_t \tag{7}$$

$$\boldsymbol{h}_t = g\left(W \boldsymbol{h}_{t-1}, U \boldsymbol{\alpha}_t\right) \tag{8}$$

$$g(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} \odot \boldsymbol{b} \tag{9}$$

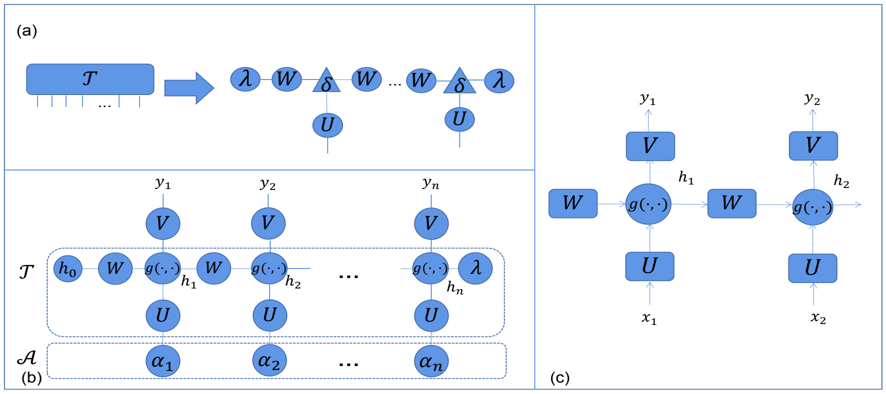

Figure 1: Recursive calculation process of TSLM. (a) represents decomposition of $\mathcal{T}$, (b) represents the process of calculating the conditional probability using $\mathcal{T}$ and $\mathcal{A}$, (c) is a general RNN architecture.

Compared with Figure 1(c), it is obvious that the deduction process is essentially the same with RNN architecture. Therefore, TSLM could be used to explain the architecture of RNN and help people understand RNN from a new perspective. The results of experiments also prove that the model makes improvement on original RNN model, thus showing its efficiency as a language model with a better theoretical foundation. Also, the success of TSLM brings tensor network into consideration. As a strong tool, it is meaningful and potential to use tensor network to improve language models in the future.

4 Conclusions and Future Work

Based on quantum many-body system, a language modeling approach based on quantum many-body wave function (QMWF-LM) was proposed at first. It shows that it is feasible to use quantum theory to represent the interactions among words with multiple meanings. What's more, a generalized language model (TSLM) was introduced inspired by QMWF-LM, adopting the same method to express the word using quantum wave function, deriving a recursive representation of the conditional probability distribution. It is promising to combine quantum theory with natural language processing tasks.

First, the efficiency of the existing models still needs to be improved. Though these models make great improvement and innovation on theory, and have solid derivation proof of every formula, it is not applicable in most of the applications due to its exponentially large dimension. And it is accesss accessible to use tensor network could be used as low-tensor approximation, which could alleviate the problem of exponentially parameters. Thus it is necessary to use tensor network to design new models based on previous theory of quantum language model.

Second, since the above models still utilize the architecture of neural network, the principle of designing networks these models remains unclear. To improve the interpretability of language model inspired by quantum theory, entanglement entropy is considered as a property of model in order to give more criteria and theoretical basis when designing the architecture of models, since the entanglement entropy in quantum many-body systems can interpret the inductive bias of networks and then guide the design of network structure and parameters for certain tasks[9, 8]. However, the dimensionality of the many-body system increases exponentially, making it difficult to calculate the entanglement entropy. We still needs to find out an effective method, which might be using tensor network since it always performs outstandingly in solving the dimensionality problem.

Overall, it is important to develop novel language representation and learning models, based on tensor network algorithms and implementation. Recently, due to the limitations of the quantum language model, some works are devoted to model natural language by tensor network [11]. However, currently these models cannot be tested on real natural language tasks, and the theoretical advantages of tensor networks, such as interpretability, have not been fully studied. Therefore, as the next step of quantum language modeling, tensor network language models need to be proposed.

References

[1] Ivano Basile and Fabio Tamburini. Towards quantum language models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1840–1849, 2017.

[2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3(6):1137–1155, 2003.

[3] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[4] William Blacoe, Elham Kashefi, and Mirella Lapata. A quantum-theoretic approach to distributional semantics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 847–857, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

[5] Peter D Bruza, Zheng Wang, and Jerome R Busemeyer. Quantum cognition: a new theoretical approach to psychology. *Trends in cognitive sciences*, 19(7):383–393, 2015.

[6] Nadav Cohen, Or Sharir, and Amnon Shashua. On the expressive power of deep learning: A tensor analysis. In *Conference on learning theory*, pages 698–728, 2016.

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

[8] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Quantum entanglement in deep learning architectures. *Physical Review Letters*, 122(6):65301, 2019.

[9] Yoav Levine, Or Sharir, and Amnon Shashua. Benefits of depth for long-term memory of recurrent networks. In *ICLR 2018 : International Conference on Learning Representations 2018*, 2018.

[10] Ding Liu, Shi-Ju Ran, Peter Wittek, Cheng Peng, Raul Blázquez García, Gang Su, and Maciej Lewenstein. Machine learning by unitary tensor network of hierarchical tree structure. *New Journal of Physics*, 21(7):073059, Jul 2019.

[11] Jacob Miller, Guillaume Rabusseau, and John Terilla. Tensor networks for language modeling. *arXiv preprint arXiv:2003.01039*, 2020.

[12] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[13] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.

[14] Alessandro Sordoni, Jian-Yun Nie, and Yoshua Bengio. 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, 2013.

[15] Edwin Stoudenmire and David J Schwab. Supervised learning with tensor networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 4799–4807. Curran Associates, Inc., 2016.

[16] Zheng-Zhi Sun, Cheng Peng, Ding Liu, Shi-Ju Ran, and Gang Su. Generative tensor network classification model for supervised machine learning. *Physical Review B*, 101(7), Feb 2020.

[17] Lipeng Zhang, Peng Zhang, Xindian Ma, Shuqin Gu, Zhan Su, and Dawei Song. A generalized language model in tensor space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7450–7458, 2019.

[18] P. Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and D. Song. End-to-end quantum-like language models with application to question answering. In *AAAI*, 2018.

[19] Peng Zhang, Zhan Su, Lipeng Zhang, Benyou Wang, and Dawei Song. 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, 2018.