A Quantum-inspired Complex-valued Representation for Encoding Sentiment Information (Student Abstract)

Guangcheng Liu, Yuexian Hou, * Shikai Song
College of Computer Science and Technology, Tianjin University, Tianjin, China
{gcliu, yxhou, sksong}@tju.edu.cn

Abstract

Recently, a Quantum Probability Drive Network (QPDN) is proposed to model different levels of semantic units by extending word embedding to complex-valued representation (CR). The extended complex-valued embeddings are still insensitive to polarity causing that they generalize badly in sentiment analysis (SA). To solve it, we propose a method of encoding sentiment information into sentiment words for SA. Attention mechanism and an auxiliary task are introduced to help learn the CR of sentiment words with the help of the sentiment lexicon. We use the amplitude part to represent the distributional information and the phase part to represent the sentimental information of the language. Experiments on three popular SA datasets show that our method is effective.

Introduction

Word embedding is a state-of-the-art technique to map words from semantic space to low-dimensional vector space based contextual information. Words are represented by tens or hundreds of dimensions of real-valued vectors. Since such training methods only consider the distributional information of the language and ignore other information such as polarity. A significant problem is different words with similar contexts will get similar vectors but correspond to opposite language polarities, such as "*good*" and "*bad*".

In Wang et al.'s work, they firstly extend word embedding to CR and model different levels of semantic units in a Semantic Hilbert Space (SHS) over the complex field. Their focus is utilizing the non-linear combined property of CR to capture implicit semantics. In this paper, we put more attention on how to use complex-valued embedding to model the polarity of language. Firstly, we use the attention mechanism to distinguish sentiment words from neutral words. Moreover, an auxiliary task is introduced to help learning the CR of sentiment words by means of the sentiment lexicon. Specifically, we use the amplitude part to represent the distributional information and the phase part to represent the sentimental information of the language. A series of systematic experiments are conducted on a Lexicon-extended Quantum Probability Driven Network (LQPDN) under the framework of quantum theory (Nielsen and Chuang 2002).

Proposed Method

With Dirac's notation, an unit vector $\vec{\mu}$ and its transpose $\vec{\mu}^T$ are denoted as ket $|u\rangle$ and bra $\langle u|$.

Quantum Probability Drive Network

Let a sentence S with N words $[w_1, w_2, \ldots, w_N]$ be an input of QPDN. Firstly we embed the input into a complex-valued matrix $[|\omega_1\rangle, |\omega_2\rangle \ldots |\omega_N\rangle]$ and $|\omega_i\rangle = \left[r_{i,1}e^{i\theta_{i,1}}, r_{i,2}e^{i\theta_{i,2}} \ldots r_{i,D}e^{i\theta_{i,D}}\right]$, where D is the dimension, r is the amplitude part and θ is the phase part of ω_i. The density matrix ρ is used to fuse word information. It is computed by outer product with each word vector and its conjugate transpose vector.

$$\rho = \sum_{i}^{N} p(w_i) |\omega_i\rangle \langle\omega_i| \tag{1}$$

where $p(w_i)$ is the weight of word with $\sum_i^N p(w_i) = 1$.

Finally a set of projectors $M = \{|m_i\rangle \langle m_i|\}_{i=1}^{K}$ are chosen to measure the mixed system ρ, which is similar to the operation of convolution kernel in CNN. The results represent the probabilities of the ρ fall onto the respective measurement operators. The training loss of QPDN $\mathcal{L}_{\text{QPDN}}$ is the cross-entropy loss between predicting labels and gold-standard labels in the dataset.

Lexicon Based Encoding Methods

A merged lexicon L helps us encode sentiment information into the complex-valued embedding. We firstly assign higher weights with sentiment words than other neutral words. For each word in S, we can define different weights as followed:

$$p(w_i) = \frac{exp(\lambda_\omega AS(w_{\text{i}}))}{\sum_{i=1}^{N} exp(\lambda_\omega AS(w_{\text{i}}))}, \tag{2}$$

$$AS(w_{\text{i}}) = |score(w_i)|. \tag{3}$$

In Eq.3, $AS(w_{\text{i}})$ is the absolute score of w_i in L. The hyper-parameter λ_ω in Eq.2 shows the degree of distinction between sentiment words and neutral words..

*Corresponding author: Yuexian Hou (yxhou@tju.edu.cn). This work is funded in part by the National Key R&D Program of China (2017YFE0111900), the National Natural Science Foundation of China (61876129) and the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 721321.

	Top 10 Similar Words	
	Real Embedding	Complex Embedding
optimistic	**pessimistic** confident upbeat hopeful pleased **disappointed** satisfied **gloomy** expect **uncertain**	upbeat confident hopeful **pessimistic** pleased expect surprised concerned **disappointed** optimism
disappointed	surprised **pleased delighted** frustrated shocked worried unhappy **satisfied** thrilled **confident**	surprised **pleased** shocked thrilled frustrated worried **delighted** unhappy embarrassed sorry

Table 1: Top-10 similar words of samples in different embeddings

In addition, we simultaneously predict polarities of those sentiment words in S as an auxiliary task when predicting the label of S. Supposing R sentiment words are chosen, we strip out their phase embeddings $[\theta_1, \theta_2, \dots, \theta_R]$. Then take them as the inputs of Sentiment Words Predicting Network (SWPN) and follow by a fully-connected layer with softmax function. The predicting sentiment score $f_w^p(w_i)$ for a word w_i is:

$$f_w^p(w_i) = softmax\left(w_2\sigma\left(w_1\theta_i + b_1\right) + b_2\right), \qquad (4)$$

where w_1, w_2, b_1 and b_2 are trainable parameters, and σ is the sigmoid function. Let $f_w^g(w_i)$ be the gold-standard sentiment score for word w_i. The target loss function for SWPN is

$$\mathcal{L}_{\text{SWPN}} = \sum_{j=1}^{C}\sum_{i=1}^{R} CE\left(\boldsymbol{f}_w^p\left(\boldsymbol{w}_{ji}\right), \boldsymbol{f}_w^g\left(\boldsymbol{w}_{ji}\right)\right), \qquad (5)$$

where w_{ji} is the i-th sentiment word in the j-th sentence, C is the batch size and $CE(\cdot)$ represents the categorical cross-entropy loss function.

Because the above two tasks share the same word embeddings, jointly learning them can take full use of sentiment lexicon from word and sentence levels. It is conducive to learn sentiment sensitive word embeddings. Finally, the target loss function of entire LQPDN is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{QPDN}} + \gamma * \mathcal{L}_{\text{SWPN}}, \qquad (6)$$

where γ is a hyper-parameter to control the degree of adjusting phase parameters.

Experimental Evaluation

We conduct experiments on MR, SST-1 and SST-2 three popular sentence-level sentiment datasets. Baselines are chosen including Multilayer Perception (MLP), Word2vec and FastText Bag-of-Words as traditional methods with shallow networks, CNN (Kim 2014) and CNN+SentiNet (Ye, Li, and Baldwin 2018) as deep learning methods with relatively complex networks, as well as our base model QPDN. Table 2 summarizes the performance of different methods.

From Table 2, we can see group 2 universally outperforms group 1. Because CNN can learn and extract features at higher level by convolution kernel. By introducing lexicon knowledge, indicators of experiments can be further improved as shown in group 2. QPDN can achieve a relatively competitive performance but not outstanding. Our proposed model outperforms all the baselines. It reflects the effectiveness of our method of encoding sentiment information into the phase part of complex-valued embedding.

Model	MR	SST-2	SST-1
MLP	77.3	79.5	39.0
Word2vec BOW	77.7	79.7	37.5
FastText BOW	78.2	80.6	40.3
CNN	78.3	83.5	43.2
CNN+SentiNet	79.6	84.2	44.0
QPDN	79.8	83.9	43.9
LQPDN	**80.8**	**84.7**	**46.6**

Table 2: The comparison results of classification accuracy over three datasets. The best scores in bold.

To test our method more intuitively, we choose two words "*optimistic*" and "*disappointed*" in SST-2 and output the top 10 similar words in L with them as shown in Table 1. Table 1 compares the Glove in 100 dimensions and the post-trained complex-valued embedding. Those words with opposite polarity are marked in bold. Since the similarity of words with the same polarity is improved while the similarity of words with the opposite polarity is reduced, complex-valued embeddings effectively distinguish these words which have the similar representation in real-valued embeddings.

Conclusion

In this paper, we adopt a new perspective from quantum theory to model sentiment polarity of word embedding. Results show that lexicon knowledge can be encoded into the phase part of complex-valued word embeddings. The new word embeddings generalize better in the overall experiments.

References

Kim, Y. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* .

Nielsen, M. A.; and Chuang, I. 2002. *Quantum computation and quantum information*. American Association of Physics Teachers.

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

Ye, Z.; Li, F.; and Baldwin, T. 2018. Encoding sentiment information into word vectors for sentiment analysis. In *Proceedings of the 27th International Conference on Computational Linguistics*, 997–1007.