Quantum-inspired Complex Word Embedding

Qiuchi Li[*]
University of Padova
Padova, Italy
`qiuchili@dei.unipd.it`

Sagar Uprety[*]
The Open University
Milton Keynes, UK
`sagar.uprety@open.ac.uk`

Benyou Wang
University of Padova
Padova, Italy
`wabyking@163.com`

Dawei Song
The Open University
Milton Keynes, UK
Beijing Institute of Technology
Beijing, China
`dawei.song@open.ac.uk`

Abstract

A challenging task for word embeddings is to capture the emergent meaning or polarity of a combination of individual words. For example, existing approaches in word embeddings will assign high probabilities to the words "Penguin" and "Fly" if they frequently co-occur, but it fails to capture the fact that they occur in an opposite sense - Penguins do not fly. We hypothesize that humans do not associate a single polarity or sentiment to each word. The word contributes to the overall polarity of a combination of words depending upon which other words it is combined with. This is analogous to the behavior of microscopic particles which exist in all possible states at the same time and interfere with each other to give rise to new states depending upon their relative phases. We make use of the Hilbert Space representation of such particles in Quantum Mechanics where we subscribe a relative phase to each word, which is a complex number, and investigate two such quantum inspired models to derive the meaning of a combination of words. The proposed models [1] achieve better performances than state-of-the-art non-quantum models on the binary sentence classification task.

1 Introduction

Word embeddings (Bengio et al., 2003; Mikolov et al., 2013; Pennington et al., 2014) are the current state of art techniques to form semantic representations of words based on their contexts. They have been successfully used in various downstream tasks such as text classification, text generation, etc. Building on word embeddings, various unsupervised (Kiros et al., 2015; Hill et al., 2016a) and supervised (Conneau et al., 2017) models for sentence embeddings have been proposed. The general idea behind word embeddings is to use word co-occurrence as the basis of semantic relationship between words. This naturally brings about the difficulty for word embedding approaches in capturing the emergent meaning of a combination of words, such as a phrase or a sentence. For example, the phrase "ivory tower" can hardly be modeled as a semantic combination of "ivory" and "tower". Or, the high frequency of occurrence of the words "Penguin" and "Fly" fails to suggest that they are negative correlated.

In the field of information retrieval (IR), various models based on the mathematical framework of Quantum Theory have been applied to capture and represent dependencies between words (Sordoni et al., 2013; Xie et al., 2015; Zhang et al., 2018), inspired by the pioneering work of Van Rijsbergen (2004). Sordoni et al. (2013) models a segment of text as a quantum mixed state, represented by a positive semi-definite matrix called density matrix in a Hilbert Space, whose non-diagonal entries entail word relations in a quantum manner(Quantum Interference). The resulting Quantum Language Model (QLM) outperforms various classical models on ad-hoc retrieval tasks. Xie et al. (2015) captures Unconditional Pure Dependence (UPD) (Hou et al., 2013) between words in a quantum way by demonstrating the equivalence relation between UPD and Quantum Entan-

glement (QE) and providing a way to incorporate UPD information into QLM, leading to improved performance over the original QLM. Zhang et al. (2018) develops a well-performing question answering (QA) system by extracting various features and learning to compare the density matrices between a question and an answer.

The successful application of quantum-inspired models onto IR tasks (Wang et al., 2016) to some extent demonstrates the non-classical nature of word dependency relations. However, all these models simplify the space of interest to be space of real vectors $\mathbb{R}^n$, with the representation of a word or a text segment being a real-valued vector or matrix, largely due to the lack of proper textual features corresponding to the imaginary part. Since quantum phenomena cannot be faithfully expressed without complex numbers, these models are theoretically limited. In a recent work, Aerts et al. (2017) presents a theoretical quantum framework for modeling a collection of documents called QWeb, in which a concept is represented as a state in a Hilbert Space, and concept combination is represented as a superposition of the concept states. Under this framework, the complex phases of each concept have a natural correspondence to the extent of interference between concepts. However, the framework has not given rise to any applicable models onto IR or NLP tasks to the authors' knowledge.

Inspired by the potential of quantum-inspired models to represent word relations, we seek to build quantum models to represent words and word combinations, and explore the use of complex numbers in the modeling process. Our model is built on top of two hypothesis: I) A word is a linear combination of latent concepts with complex weights. II) A combination of words is viewed as a complex combination of word states, either a superposition state or a mixed state. The first hypothesis agrees with QWeb, but here we concretize a concept in QWeb to be a word. The second hypothesis is an extension of both QWeb and the work by Zhang et al. (2018), because QWeb restricts a combination of concepts to be a superposition state while the work by Zhang et al. (2018) assumes that a sentence is a complex mixture of word projectors.

This study sets foot in sentence-level analysis, and treats a sentence as a combination of words. We intend to model a word as a quantum state containing two parts: amplitudes and complex phases, and expect to capture the low-level word co-occurrence information by the amplitudes, while using the phases to represent the emergent meaning or polarity when a word is combined with other words. We investigate on two models to represent the combination of words, either as a superposition of word states or as a mixture of word projectors. The effectiveness of the two models are evaluated on 5 benchmarking binary sentence classification datasets, and the results show that the mixture model outperforms state-of-the-art word embedding approaches.

The motivation behind this paper stems from an analogy with Quantum Physics. Consider the phrase "Penguins fly". If we model it along the lines of the famous double slit experiment in Quantum Physics, the two slits corresponds to human interpretation of words "Penguins" and "Fly" (Verb sense of Fly). When only one slit is open at a time, the waves corresponding to the individual word will go through the slit and register onto the screen. The screen is made of a set of polarity detectors judging opinion or sentiment polarities.

In Figure 1.a the human mind sees the word 'Penguins' alone and detects it as a neutral word with a very high probability. This is analogous to the double slit experiment with one slit open. The same is the case for the word 'Fly' considered in isolation. By classical logic, when the two words are taken together as a phrase 'Penguins fly', the human mind should assign a high probability of it being neutral again. However, we know that it is a false statement (Figure 1.c).

Different from classical representation, this study hypothesizes that the combination of words can be viewed as a superposition or complex mixture of quantum entities which gives rise to a new state. In this way, the emerging meaning or polarity of a combination of words will manifest in the interference between words, and be captured inherently in the density matrix representation. For example, two or more words having a neutral sense individually may combine to give a negative sense, just like the case in the analogy given above.

2 Hilbert Space Representation of Words and Sentences

The section introduces the proposed quantum framework for representing words and sentences.

(a) Figure 1.a

(b) Figure 1.b

(c) Figure 1.c
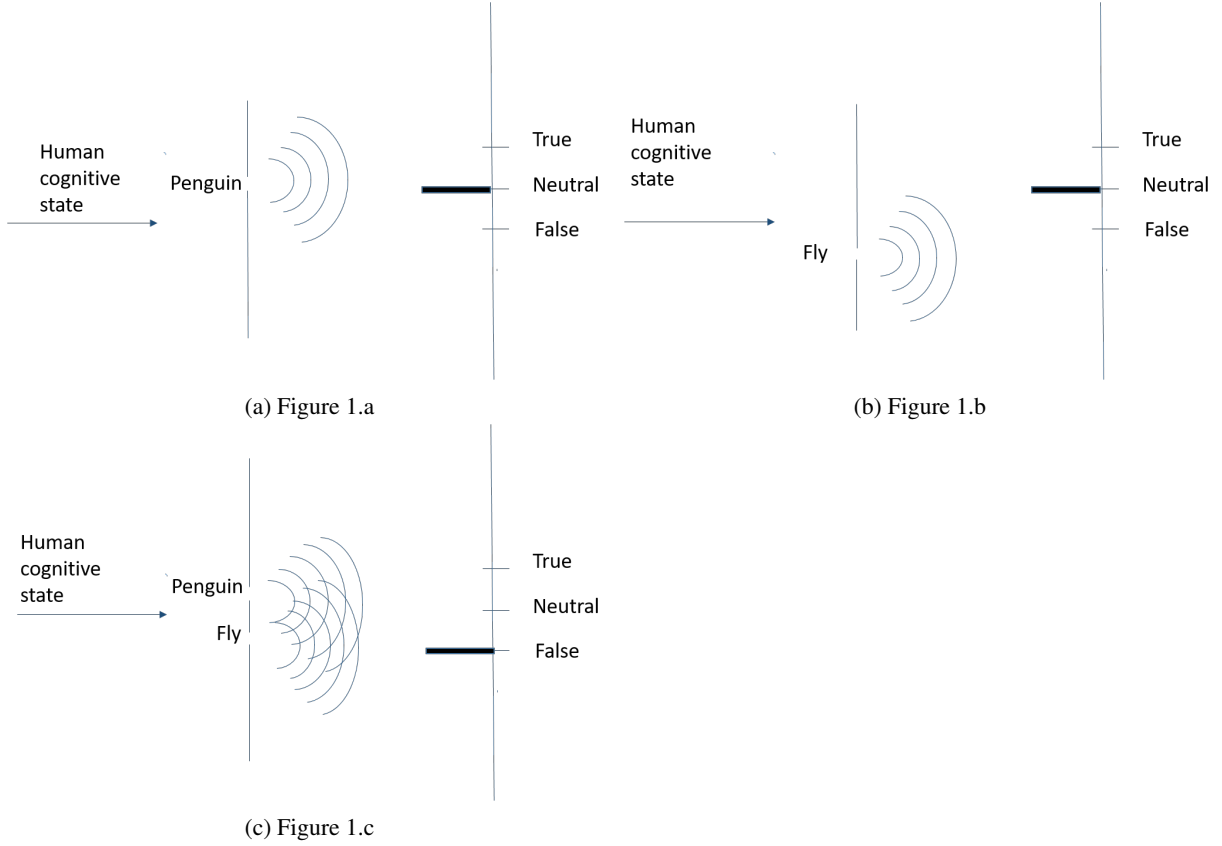
Our research scope is currently limited to sentence and word level analysis. However, our proposed model is potentially capable of representing higher-level concepts such as paragraphs and documents, which we will investigate in the future. To be consistent with the quantum framework, we use Dirac notations, in which a unit vector $\vec{\mu}$ and its transpose $\vec{\mu}^T$ are denoted as ket $|u\rangle$ and bra $\langle u|$ respectively.

Suppose there are n independent latent concepts in the text collection, we then model words and sentences as quantum concepts defined on an n-dimensional Hilbert Space $\mathbb{H}^n$, where latent concepts form a set of pure orthonormal states of the space. Using Dirac notations, the concepts are denoted as $\{|C_i\rangle\}_{i=1}^n$. Intuitively, latent concepts correspond to the contexts in which words are used.

Each word t is modeled as a superposition state (Nielsen and Chuang, 2011) in the n-dimensional Hilbert Space $\mathbb{H}^n$. Equivalently, it can be viewed as a linear combination of $\{|C_i\rangle\}_{i=1}^n$ with complex weights, i.e. $|t\rangle = \sum_{k=1}^n e^{i\theta_k} w_k |C_k\rangle$, in which $\{w_i\}_{i=1}^n$ are real-valued amplitudes with $w_i > 0$ and $\sum_{i=1}^n w_i^2 = 1$, and $\theta_i \in [-\pi, \pi], i = 1, 2, ..., n$ are the corresponding complex phases. This representation can

be seen as a generalization of previous word embedding approaches (Bengio et al., 2003; Mikolov et al., 2013; Pennington et al., 2014) in that it can be regarded as a complex embedding with unitary length of word vectors. A word has many different contexts associated with it. For example, 'Penguin' is associated with 'Bird', 'Antarctica', 'Snow', etc. When a quantum particle(e.g. electron) is said to be in a superposition state, it exists in a new state(e.g. position) of all of its possible outcomes(at all positions) at the same time. A particular outcome is observed upon measurement. Similarly a word exists in all of its contexts at the same time and depending upon its interaction with other words in a combination, a particular context is materialized. Note that because of reduced dimensionality, the contexts are latent concepts.

A sentence is a non-classical combination of words. Since each word is a superposition of latent concepts, a sentence s is also a non-classical combination of latent concepts $\{|C_i\rangle\}_{i=1}^n$. It is represented by a n by n density matrix ρ which is positive semi-definite with unitary trace: $\rho \geq 0$, $\text{Tr}(\rho) = 1$. The real diagonal values of ρ reflects the strength of concepts in the sentence, whereas the non-diagonal values encodes correlations between concepts in a quantum manner. The density ma-

trix can be computed from the word states either directly or through a training strategy.

Our proposed approach is related to but largely differs from Sordoni et al. (2013) and Zhang et al. (2018). Sordoni et al. (2013) models queries and documents as density matrices and provides a training method for constructing density matrices from texts. Zhang et al. (2018) directly computes the density matrix of a sentence and put it into an end-to-end neural network for handling the Question Answering (QA) task. Both works view a segment of texts as a mixed state (Nielsen and Chuang, 2011) and use real-valued density matrix as a representation. Our study also directly computes the sentence representation from the word superposition states. However, different from both works, our study explores on treating a sentence as either a strictly mixed state or a superposition state. In either case, it can be represented as a complex density matrix with complex values for non-diagonal entries.

On top of the obtained sentence representation, different quantum operations can be applied to achieve a particular NLP target at hand. For sentence classification tasks, one can perform projective measurements onto the sentence representation to determine the sentiment polarity; for sentence text similarity task, the amplitude of the inner product between a sentence pair may provide evidence for judging to what extent they are similar to each other. Projective measurements and inner products are methods to compute probabilities in Quantum Theory (Nielsen and Chuang, 2011).

3 Complex Embedding Network for Text Classification

In this paper, we build a complex embedding network for text classification on the basis of Hilbert Space representation for words and sentences. The end-to-end network accepts a sentence sequence as input and computes its classification label in the procedure shown by Figure 2:

The input one-hot sequence is passed through an embedding layer with a complex valued lookup table, which maps each word into a complex vector representing its superposition state, resulting in a sequence of complex embedding vectors. Then the density matrix of the sentence is computed from the complex embedding vectors. Finally, a square projection matrix takes control of the measurement. For any sentence state ρ, the mea-

surement probability is computed through Born's rule (Born, 1926):

$$p = Tr(P\rho) \tag{1}$$

Where P is a projection matrix satisfying $P^2 = P, P = P^T$. The value of p determines the class of this sentence. The lookup table determining the complex embedding for each word is learned by feeding the network with a sufficient number of training data.

The crucial step of the process falls on how to compute the sentence density matrix from the sequence of complex word embeddings. As no previous research has attempted to build complex networks for text classification task, we investigate on two approaches for this step:

I) A sentence is viewed as a linear combination of all word vectors in the sentence, i.e. $|S\rangle = \frac{\sum_{l=1}^{m} \lambda_l |t_l\rangle}{||\sum_{l=1}^{m} \lambda_l |t_l\rangle||_2}$, with $\sum_{l=1}^{m} \lambda_l = 1$. Here λ_ls are real-valued weights indicating the relative degree of importance for each word in the sentence, and the state is divided by its 2-norm in order to guarantee it is a legal quantum state (i.e.,with unit length). The sentence is then a pure superposition state and the density matrix can be computed simply as $\rho = |S\rangle\langle S|$.

II) A sentence is viewed as a classical mixture of the word states in the sentence, i.e. $\rho = \sum_{l=1}^{m} \lambda_l |t_l\rangle\langle t_l|$, with $\sum_{l=1}^{m} \lambda_l = 1$. Here $|t_l\rangle\langle t_l|$ is the density matrix representing the superposition state of a word t_l. This equation guarantees the obtained ρ is a legal density matrix without any further normalization.

The constructed density matrix representing a sentence has real values for diagonal entries and non-zero complex values for non-diagonal entries. Intuitively, the diagonal entries tell us something about the distribution of latent concepts in the sentence, whereas the non-diagonal values entail information regarding the emergent meanings. Consider a very simple example where the complex phases represent positive, neutral or negative senses. Independently, both the words "Penguin" and "Fly" have neutral sense, $\theta_P = \theta_F = 0$. When they are combined together in a sentence, then sentence density matrix has a negative-phased complex value in the entry corresponding to them, i.e. $\theta_{PF} < 0$. Therefore, the combination of these two words will have a negative complex phase, implying the negative sense "Penguins cannot fly".

Figure 2: The process diagram of the proposed complex embedding network. $|V|$ is the vocabulary size, n is the embedding dimension, m is the maximum length of a sentence
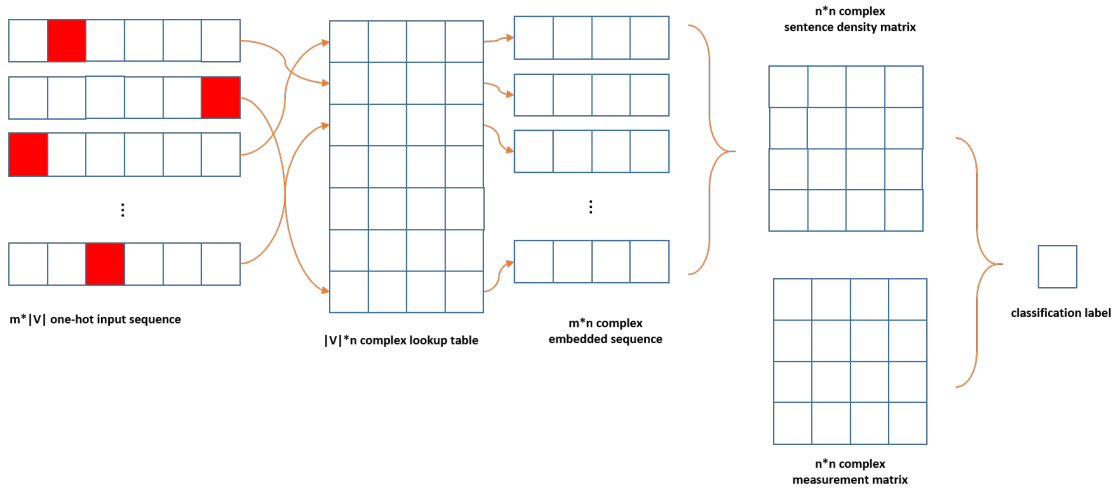
In practice, the connections between words are much more complicated, but we believe that by feeding the above-mentioned models with enough data, the constructed density matrix will be able to effectively capture and represent the emergent meanings of sentences.

The above-mentioned approaches lead to two different models, resulting in different embeddings learned from the same training data. Hence we name them as complex embedding superposition (CE-Sup) network and complex embedding mixture (CE-Mix) network respectively. For sake of simplicity, we assign equal importance of each word in the sentence representation in both models, i.e. $\lambda_l = \frac{1}{m}, l = 1, 2, ..., m$. In a relevant research, Zhang et al. (2018) learns the values of λ_ls in the training framework, while enforcing the word embeddings $|t_l\rangle$s to be fixed. By fixing λ_ls and learning $|t_l\rangle$s from the data, this paper is essentially aiming at obtaining better representation of each word from the training data, whereas Zhang et al.'s work directly takes existing word vectors trained from external corpus. It would be interesting to see what a co-training of $|t_l\rangle$s and λ_ls will bring about in future works.

4 Experimental Setup

The experiments are conducted on five benchmarking datasets for binary text classification: Customer Review dataset (CR) (Hu and Liu, 2014), Opinion polarity dataset (MPQA) (Wiebe et al., 2005), Sentence Subjectivity dataset (SUBJ) (Pang and Lee, 2005), Movie Review dataset (MR) (Pang and Lee, 2005), and Stanford Sentiment Treebank (SST) dataset [2]. The statistics for the datasets are shown in Table 1.

Table 1: Dataset Statistics

Dataset	#Count	Task	Classes
CR	4k	product reviews	pos/neg
MPQA	11k	opinion polarity	pos/neg
SUBJ	10k	subjectivity	subj/obj
MR	11k	movie reviews	pos/neg
SST	70k	movie reviews	pos/neg

In this paper, we compare the classification accuracy of our proposed Complex Embedding Superposition (CE-Sup) network and Complex Embedding Mixture (CE-Mix) network with three existing unsupervised representation training models, Unigram-TFIDF and fastText Bag-of-Words (BOW), as well as two existing supervised representation training models, namely CaptionRep BOW (Hill et al., 2016b) and DictRep BOW (Hill et al., 2016c). We directly take the performances of these systems on the 5 datasets from existing works. Since the performances for CaptionRep and DictRep are not available on SST, we use the performance of another model called Paragraph-Phrase (Bansal and Livescu, 2016). For a fair comparison, we also implement an end-to-end supervised real embedding network (Real-Embed),

where each word is mapped to a real-valued vector in the embedding layer, based on which the sentence representation is obtained by averaging the embedding vectors for all words in the sentence, and a fully connected layer maps the sentence vector to the classification label. CE-mixture, CE-Superposition and Real-Embed are trained and tested in a completely identical process.

For the construction of training, validation and test data, they are readily available for SST dataset, and for the other four datasets we randomly split the whole data into 8:1:1 for training, validation and test data respectively. The embedding dimension is set to be 100. We use batch training with batch size being 32 for SST and 16 for the other datasets. We adopt Adam as the optimizer and use the default parameters for Adam in Keras [3].

The experiments are implemented in Keras and Tensorflow [4] under Python 3.6.4. The experiment is run on a desktop with NVidia Quadro M4000 and 16GB RAM.

5   Results and Discussion

In this study, we seek to answer the following two research questions:

RQ1. Do the proposed quantum-inspired complex embedding models outperform state-of-the art non-quantum approaches?

RQ2. Out of the two proposed model in this study, which one performs better?

Table 2 presents the classification accuracy values of all models experimented in this paper, where the bold values indicate the best-performing models for each dataset. It can be clearly seen from the table that CE-Mix is the best-performing model, because it occupies the highest accuracy value on 4 out of 5 benchmarking datasets, and on the remaining dataset it performs only slightly worse than the best-performed model.

In order to make the results more convincing, we also conduct two-tailed p-tests on the performances. The hypotheses are:

H0. There is no difference between two groups of performances on a particular dataset.

H1. There is a difference between two groups of performances on a particular dataset.

We use the threshold 0.05 to accept or reject the null hypothesis: when the obtained p-value < 0.05, the null hypothesis is rejected; when p-value > 0.05, the null hypothesis is accepted.

Regarding RQ1, it can be observed that CE-Sup and CE-Mix achieves consistently higher or comparable accuracy than non-quantum models under experiment. It illustrates the superiority of complex embedding network over traditional language model (Unigram-TFIDF) (p-value < 0.05 on all datasets, rejecting the null hypothesis, and so forth), unsupervised embeddings trained from external corpus (word2vec, fastText) (p-value < 0.05 on all datasets except MPQA), as well as supervised embedding methods (CaptionRep, DictRep and Paragram-Phrase) (p-value$<$ on all datasets except MPQA). The fair comparison with real embedding network (p-value < 0.05 on all datasets) confirms the superiority of complex embedding over real embedding techniques.

Regarding RQ2, out of the two complex embedding models proposed in this study, CE-Mix performs consistently but insignificantly (p-value > 0.05) better than CE-Sup in all datasets. Even though it is yet a fully convincing evidence, this result provides us with some intuition that it seems better to model a sentence as a classical mixture of word projectors rather than as a superposition state of latent concepts. For future work we will evaluate the performances of these two models on other datasets as well as other tasks to reach a more solid conclusion.

6   Conclusion and Future Work

This paper attempts to address the challenge of representing the combinatory meaning of words for word embedding. The successful applications of quantum-based models in IR tasks inspires us to construct Hilbert Space representation of words and sentences, and explore to build two quantum models for solving sentence classification task. The experimental result on five benchmarking datasets demonstrates their effectiveness.

This work contributes to the fields of both word embeddings and quantum-inspired IR. On the one hand, our work can be interpreted as an improved embedding approach, which tackles the challenge

Table 2: Experimental Results in percentage(%). The best performed value for each dataset is in bold.

Model	CR	MPQA	MR	SST	SUBJ
Unigram-TFIDF	79.2	82.4	73.7	-	90.3
word2vec BOW	79.8	**88.3**	77.7	79.7	90.9
fastText BOW	78.9	87.4	76.5	78.8	91.6
CaptionRep BOW	69.3	70.8	61.9	-	77.4
DictRep BOW	78.7	87.2	76.7	-	90.7
Paragram-Phrase	-	-	-	79.7	-
Real-Embed	77.5	84.7	77.0	80.0	92.0
CE-Sup	80.0	85.7	78.4	82.6	92.6
CE-Mix	**81.1**	86.6	**79.8**	**83.3**	**92.8**

of capturing the emergent meaning of a combination of words. On the other hand, this can be viewed as a pioneering study on quantum-inspired language models with complex numbers, and also an trial effort to adopt the theoretical QWeb framework onto an application context.

For future work, it is necessary to conduct a more comprehensive evaluation of the proposed models, either by evaluating on more datasets or by evaluating the qualities of the trained complex embeddings. We are also looking forward to seek additional ways to model a sentence based on the word states, and the application of the models onto other NLP tasks.

ACKNOWLEDGEMENT

References

Diederik Aerts, Jonito Aerts Arguelles, Lester Beltran, Lyneth Beltran, Isaac Distrito, Massimiliano Sassoli de Bianchi, Sandro Sozzo, and Tomas Veloz. 2017. Towards a Quantum World Wide Web. *arXiv:1703.06642 [quant-ph]* ArXiv: 1703.06642. http://arxiv.org/abs/1703.06642.

John Wieting Mohit Bansal and Kevin Gimpel Karen Livescu. 2016. TOWARDS UNIVERSAL PARAPHRASTIC SENTENCE EMBEDDINGS page 19.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *J. Mach. Learn. Res.* 3:1137–1155. http://dl.acm.org/citation.cfm?id=944919.944966.

Max Born. 1926. Zur Quantenmechanik der Sto\s svorgnge. *Zeitschrift fr Physik* 37(12):863–867. https://doi.org/10.1007/BF01397477.

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 670–680. http://aclweb.org/anthology/D17-1070.

Felix Hill, Kyunghyun Cho, and Anna Korhonen. 2016a. Learning distributed representations of sentences from unlabelled data. *CoRR* abs/1602.03483. http://arxiv.org/abs/1602.03483.

Felix Hill, Kyunghyun Cho, and Anna Korhonen. 2016b. Learning Distributed Representations of Sentences from Unlabelled Data. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, pages 1367–1377. http://www.aclweb.org/anthology/N16-1162.

Felix Hill, KyungHyun Cho, Anna Korhonen, and Yoshua Bengio. 2016c. Learning to Understand Phrases by Embedding the Dictionary. *Transactions of the Association for Computational Linguistics* 4:17–30. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/711.

Yuexian Hou, Xiaozhao Zhao, Dawei Song, and Wenjie Li. 2013. Mining Pure High-order Word Associations via Information Geometry for Information Retrieval. *ACM Trans. Inf. Syst.* 31(3):12:1–12:32. https://doi.org/10.1145/2493175.2493177.

Minqing Hu and Bing Liu. 2014. Mining and Summarizing Customer Reviews page 10.

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. Skip-thought

vectors. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. MIT Press, Cambridge, MA, USA, NIPS'15, pages 3294–3302. http://dl.acm.org/citation.cfm?id=2969442.2969607.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. Curran Associates Inc., USA, NIPS'13, pages 3111–3119. http://dl.acm.org/citation.cfm?id=2999792.2999959.

Michael A. Nielsen and Isaac L. Chuang. 2011. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition.

Bo Pang and Lillian Lee. 2005. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. Association for Computational Linguistics, pages 115–124. https://doi.org/10.3115/1219840.1219855.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*. volume 14, pages 1532–1543.

Alessandro Sordoni, 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*. ACM, New York, NY, USA, SIGIR '13, pages 653–662. https://doi.org/10.1145/2484028.2484098.

Cornelis Joost Van Rijsbergen. 2004. *The geometry of information retrieval*. Cambridge University Press.

Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. 2016. Exploration of quantum interference in document relevance judgement discrepancy. *Entropy* 18(4):144.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39(2-3):165–210. https://doi.org/10.1007/s10579-005-7880-9.

Mengjiao Xie, Yuexian Hou, Peng Zhang, Jingfei Li, Wenjie Li, and Dawei Song. 2015. Modeling Quantum Entanglements in Quantum Language Models. In *Proceedings of the 24th International Conference on Artificial Intelligence*. AAAI Press, Buenos Aires, Argentina, IJCAI'15, pages 1362–1368. http://dl.acm.org/citation.cfm?id=2832415.2832439.

Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. 2018. End-to-End Quantum-like Language Models with Application to Question Answering .