Semantic Similarities using Classical Embeddings in Quantum NLP

Damir Cavar
NLP-Lab
Indiana University at Bloomington
Bloomington, IN, USA
dcavar@iu.edu

Chi Zhang
NLP-Lab
Indiana University at Bloomington
Bloomington, IN, USA
czh4@iu.edu

***Abstract*—We demonstrate how word and text embeddings as n-dimensional real vectors used in classical Natural Language Processing (NLP)/AI applications can be mapped to quantum states or quantum embedding representations in Quantum Applications using NLP and AI. Models like fastText [1], [2], GloVe [3], Numberbatch [4], or BERT [5] are common NLP embedding vectors or models that encode semantic properties of words or text fragments. We use these models to evaluate mapping properties and compression rate, as well as information preservation in quantum embeddings. For mapping embedding vectors to quantum states, we use encoding strategies, such as Amplitude Encoding. The encoding strategies allow us to map large dense vectors from the embedding models to compact quantum states using different compression ratios. The compression with Amplitude Encoding can be 2^n to N, resulting in a mapping of a 1,024-dimensional vector of reals in the classical environment to a 10-qubit state. The goal of this work is to evaluate these strategies with respect to their compression ratio and measure the preservation of semantic information using similarity scores. We show that the resulting quantum embeddings based on mapped classical computing embeddings exhibit the same relational properties and that there is no significant loss of semantic information in the conversion from classical n-dimensional real vector embeddings to qubit states. We conclude that the experimental results allow us to quantum compute semantic similarities of words or text, reusing existing and freely available embedding models from classical NLP/AI computing.**

***Index Terms*—Word embeddings, Semantic Similarity, Quantum Embeddings, Natural Language Processing**

I. Introduction

Vector representation of linguistic units, for example, words, sentences, or text in classical Natural Language Processing (NLP) and AI applications, is essential for search algorithms, neural network architectures, and machine learning algorithms. Instead of using a specific vector length of arbitrary real values, such vectors are trained using Distributional Semantics [6]. That is, the vectorized representations of words in the form of dense vectors should contain semantic properties that can be measured in terms of similarity. Semantically related words are closer to each other than semantically unrelated words by means of Linear Algebra. The similarity in these models can be measured using, e.g., *cosine similarity* or *Euclidean distance* between dense vectors (i.e., word embeddings).

Rath and Date (2023) [7] demonstrate that encoding methods that map dense classical vectors to quantum embeddings, e.g., basis encoding, angle encoding, or amplitude encoding used in popular Machine Learning (ML) algorithms, can be used successfully used in Quantum Computing (QC), and the resulting quantum embeddings can in fact ”contribute to improved classification accuracy and F1 scores” in the relevant applications.

While these encoding methods have been reported to be adequate for Quantum ML applications [7], to the best of our knowledge, we do not have found experimental results measuring the information loss related to semantic similarity in the different encoding methods from classical embedding vectors to quantum embeddings. The work here is focusing on exactly that.

II. Quantum Word Similarities: Methodology

Using classical word embedding models we generate vector representations for individual words or word sequences. We encode these vectors as quantum embeddings using various methods as for example basis encoding or amplitude encoding. We then measure the quantum similarity between two quantum embeddings.

Our approach to defining the quantum similarity between two words is via the SWAP test [8]. The SWAP test can be performed between any two circuits S and T of the same number of qubits and is a way to measure the difference between S and T. Suppose the qubits in S are named $s_0, s_1, \dots, s_{n-1}$ and those in T are $t_0, t_1, \dots, t_{n-1}$ with an ancillary qubit q_0, the SWAP test performs first the Hadamard gate then the controlled SWAP gate from q_0 to s_i and $t_i, i = 0, \dots, n-1$ and again the Hadamard gate. It then measures the value of q_0. The circuit for the SWAP tests is plotted in Fig.1.

Suppose we use amplitude encoding from [9] to encode 2 $N = 2^n$ real vectors into the circuits S and T. If we write each state $|s_i\rangle$ as

$$|s_i\rangle = \frac{1}{\|s_i\|} \sum_{j=0}^{N-1} s_{ij} |j\rangle$$

979-8-3315-4137-8/24/$31.00 ©2024 IEEE
DOI 10.1109/QCE60285.2024.10350

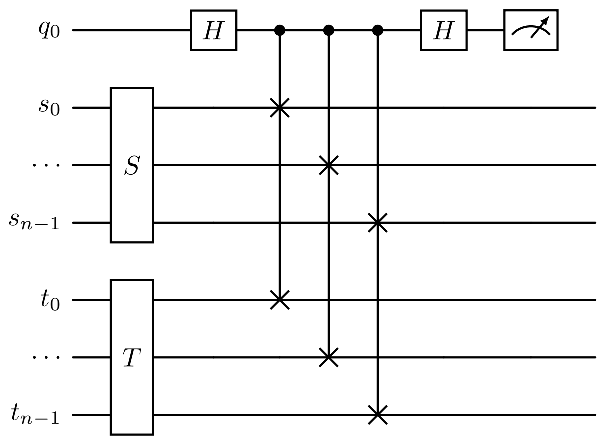

Fig. 1. Quantum circuit for the SWAP test between two circuits S and T

where $|j\rangle\, j = 0, \dots N-1$ is the standard basis of the Hilbert space $\mathbb{C}^{\otimes n} \simeq (\mathbb{R}^2)^{\otimes n} \simeq \mathbb{R}^{2^n} = \mathbb{R}^N$, we can write the quantum state $|S\rangle$ of the circuit S as

$$|S\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} |i\rangle \otimes |s_i\rangle \otimes |b_i\rangle$$

where $|b_i\rangle$ is the eigenstate of the Pauli matrix σ_z of eigenvalue $y_i \in \{0,1\}$. After a straightforward calculation [10], we know the probability of measuring q_0 having outcome 1 is

$$\begin{aligned} P(1; S, T) &= \frac{1}{4}(1 - \frac{1}{N\sqrt{2}} \sum_{i=0}^{N-1} y_i \langle s_i | t_i \rangle) \\ &= \frac{1}{4}(1 - \frac{1}{N\sqrt{2}} \sum_{i=0}^{N-1} y_i \cos(s_i, t_i)). \end{aligned} \tag{1}$$

Now suppose we have two words `word1` and `word2`, and suppose their vector representations in some model, *e.g.* `fastText` are `vec(word1)` and `vec(word2)`. After applying amplitude encoding to `vec(word1)` and `vec(word2)`, we denote the corresponding quantum circuits `circ(word1)` and `circ(word2)`, we define the SWAP distance `swap_dist(word1, word2)` between `word1` and `word2` to be

$$\begin{aligned} &\texttt{swap\_dist(word1, word2)} \\ &= P(1; \texttt{circ(word1), circ(word2)})' \end{aligned} \tag{2}$$

which is the probability of measuring 1 from q_0 with $S = \texttt{circ(word1)}$ and $T = \texttt{circ(word2)}$

III. Data and Results

We used the following models to select word pairs and measure classical similarities (Cosine Similarity) using the model embeddings. All models except Numberbatch [4] were English word lists, but many of those contained non-English words and non-words or non-linguistic expressions. The Numberbatch model is multilingual that is trained on the ConceptNet knowledge graph. BERT is not a model that consists of word and vector pairs, but rather a transformer [11] that is a deep learning [12] model that generates a vector representation for some input text.

- fastText, 300-dimensional word vectors, 2.5 mil. words
- GloVe, 840 billion tokens, 300-dimensional word vectors, 2.1 mil. words
- Numberbatch, 300-dimensional vectors, 516,783 words
- BERT, 768-dimensional word vectors

For the experiments, we selected 4400 randomly picked word pairs that exist in all models. We computed the Cosine Similarity and the Quantum Word Similarity for all pairs using the vector representations from the four models listed above.

In addition to the randomly selected words, we created a word list of 100 words that belong to different semantic topics or fields, for example:

- *apple tree grape vine*
- *car truck bus bike motorcycle*
- *doctor nurse surgeon dentist vet*

Measuring the Correlation Coefficient over 4400 word pairs using Cosine Similarities and Quantum similarities we achieve a score of almost 0.90 on average for the pre-computed vector models, indicating that the information loss is minimal and insignificant.

The code and data for this poster are available at https://github.com/dcavar/q-embeddings-QCE24.

References

[1] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[2] A. Joulin, P. Bojanowski, T. Mikolov, H. Jégou, and E. Grave, "Loss in translation: Learning bilingual word mapping with a retrieval criterion," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.

[3] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162

[4] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," 2017, pp. 4444–4451. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 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)*, J. B. et al., Ed. Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[6] A. Lenci and M. Sahlgren, "Distributional semantics," 2023.

[7] M. Rath and H. Date, "Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy," *ArXiv*, vol. abs/2311.10375, 2023.

[8] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf, "Quantum Fingerprinting," *Physical Review Letters*, vol. 87, no. 16, p. 167902, Sep. 2001.

[9] I. F. Araujo, D. K. Park, F. Petruccione, and A. J. Da Silva, "A divide-and-conquer algorithm for quantum state preparation," *Scientific Reports*, vol. 11, no. 1, p. 6329, Mar. 2021.

[10] D. Pastorello and E. Blanzieri, "A quantum binary classifier based on cosine similarity," in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Oct. 2021, pp. 477–478.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.