Instructions for ACL 2023 Proceedings

Anonymous ACL submission

Abstract

This is abstract

1 Introduction

2 Previous studies

3 Distance vectors of derivation class

This section utilizes the distance vectors between derived sentence embeddings and the seed sentence embedding to predict the derived embedding for another seed. We hypothesize that the distance vectors for each derivation class share a similarity and are able to predict the derivation embeddings given the seed embedding. For instance, for the derivation class 'future', we assume that the following property holds: given two seed sentences $Seed_i$ and $seed_j$, and one future sentence, $future_i$, we can predict $future_j$.

$$future_j = future_i - seed_i + seed_j \quad (1)$$

To test this hypothesis, we calculate the cosine similarity between the true sentence embeddings and predicted sentence embeddings using Equation 1. We use 80% of the data to extract and average the distance vectors ($sent_i$-$seed_i$) for each derivation class. We then add these distance vectors to the seed vector for the remaining 20% of the data and compare the cosine similarity between the predicted and true embeddings.

The results are presented in Table 1. Most of the derivation classes achieved a cosine similarity score above 0.8 compared with true embeddings, indicating that the predicted vectors closely approximate the true vectors in space, and the distance vectors of derived sentences and seed sentences are relatively stable for each derivation class.

However, the 'generalization' class showed the lowest score, which could be attributed to the fact that each seed contains multiple degrees of transformation for the 'generalization' class. Some derived sentences are more generalized, while others are less so, and averaging 80% of them may not be sufficient to capture the features of the remaining 20% of the data.

class	cosine-sim
possibility	0.936635
past	0.930636
future	0.923555
different meaning	0.913798
nonsense	0.901452
formal sentence	0.882741
minimal change	0.866246
ban	0.845272
paraphrase	0.81659
nonstandard sentence	0.812346
simple sentence	0.807523
opposite meaning	0.748853
generalization	0.697715

Table 1: Cosine Similarity of predicted and true derivation sentence embeddings

4 Methodology of Embedding Evaluation

In addition to visualizing the embeddings of derivation sentences, we employ several approaches to test whether our trained sentence embeddings can accurately characterize the derivation sentence types.

Firstly, we use the Calinski-Harabasz Index, an internal cluster validation test, to pairwise test the derivation clusters and evaluate how well a pair of derivation clusters are separated in space.

Secondly, an unsupervised Gaussian Mixture model is employed to cluster each pair of derivation classes and compute the accuracy of unsupervised cluster labels and with the true labels.

Lastly, we employ several supervised approaches to classify the embeddings and evaluate the extent to which the derivation classes can be captured by the trained classifiers. The accuracy scores are compared among classifiers including Multilayer Perceptron, Decision tree and Support Vector Machine.

5 Evaluating Sentence Embeddings with Gaussian Mixture

This section presents an unsupervised approach for evaluating the separability of sentence embeddings. We measure label separability pairwisely using a Gaussian Mixture model and calculate the F-score of the unsupervised clustered labels with the true labels.

The Gaussian Mixture model is a probabilistic model that assumes that each cluster follows a Gaussian (or normal) distribution and estimates the weight of the density function for each cluster (Reynolds et al., 2009; Singh et al., 2010). We assume that the sentence vectors of two distinct classes should achieve high accuracy with Gaussian Mixture if they are displayed in a Gaussian distribution in space and are separable from each other.

However, there is a potential limitation to using this method. The separability and accuracy score may be underestimated if two clusters are not normally distributed, as illustrated in Figure 1.

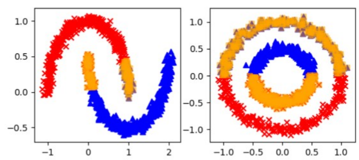

Figure 1: Cases that Gaussian Mixture Underestimate the Separability of Two Clusters

Our evaluation results show that out of 78 pairs of derivation classes, 7 pairs achieved an accuracy score above 90%, 24 pairs above 80%, and 47 pairs above 60%, as depicted in Figure 2.

In particular, the class 'ban' demonstrates good separability with many other classes, achieving an accuracy score of 0.982 with 'past', 0.980 with 'formal sentence', 0.942 with 'minimal change', and 0.937 with 'future', among other pairs.

Additionally, our evaluation results reveal that tenses are generally well-separated, with an accuracy score of 0.90 for 'past' and future' classes, and an accuracy score of 0.924 for 'simple sentence' and 'future' classes.

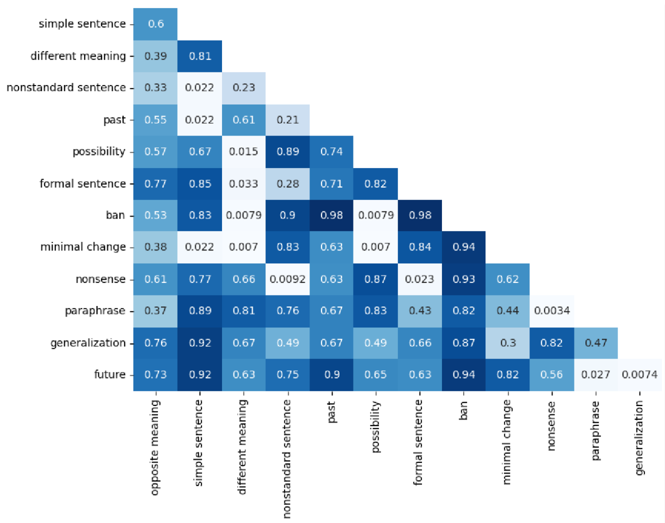

Figure 2: Accuracy of Measured Separability with Gaussian Mixture

References

Douglas A Reynolds et al. 2009. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663).

Ravindra Singh, Bikash C. Pal, and Rabih A. Jabr. 2010. Statistical representation of distribution system loads using gaussian mixture model. *IEEE Transactions on Power Systems*, 25(1):29–37.