Cognitive Sciences and AI

Assignment 3

Report

Hariharan Kalimuthu
2020115015

Note on the two metrics used for comparing the word embeddings:

The choice between using 2v2 accuracy score and Pearson correlation as a metric for comparing word embeddings depends on the specific task and the research question being addressed.
The 2v2 accuracy score is a binary classification metric that measures the ability of the word embeddings to distinguish between two sets of word pairs, with one set representing the desired semantic relationship (target words) and the other set representing a different semantic relationship (distractor words). It is a straightforward and interpretable metric that provides a clear indication of how well the word embeddings can perform a specific task, such as word similarity or analogy.

On the other hand, Pearson correlation is a continuous metric that measures the linear relationship between two sets of variables. It captures the degree of similarity between the two sets of cosine similarities and can be used to evaluate the quality of word embeddings in capturing various semantic relationships between word pairs, including those that are not explicitly represented in the set of target and distractor word pairs.

In general, 2v2 accuracy score is preferred for tasks that involve binary classification, such as word similarity and analogy. Pearson correlation, on the other hand, is preferred for evaluating the overall quality of the word embeddings across a range of semantic relationships and tasks.

Therefore, both metrics are useful for evaluating word embeddings, and the choice between them should be based on the specific research question and task at hand.

Comparison of the three embeddings:

When building a brain encoding and decoding model using fMRI data with sentence stimuli, the choice of word embeddings (e.g., CLS, Pooled, GloVe) can impact the performance of the model.
In general, CLS and Pooled embeddings are often used in natural language processing (NLP) tasks that involve sentence-level predictions, such as text classification or sentiment analysis. These embeddings are specifically trained to capture the overall meaning or sentiment of a sentence and may be more effective at encoding and decoding sentence-level information in the fMRI data. Therefore, it is possible that CLS or Pooled embeddings would perform better in a brain encoding and decoding model that uses sentence stimuli.

GloVe embeddings, on the other hand, are trained on large-scale co-occurrence data and may be more effective at capturing the semantic relationships between individual words within a sentence. While GloVe embeddings can also be used for sentence-level tasks, they may be less effective at encoding and decoding sentence-level information in the fMRI data.

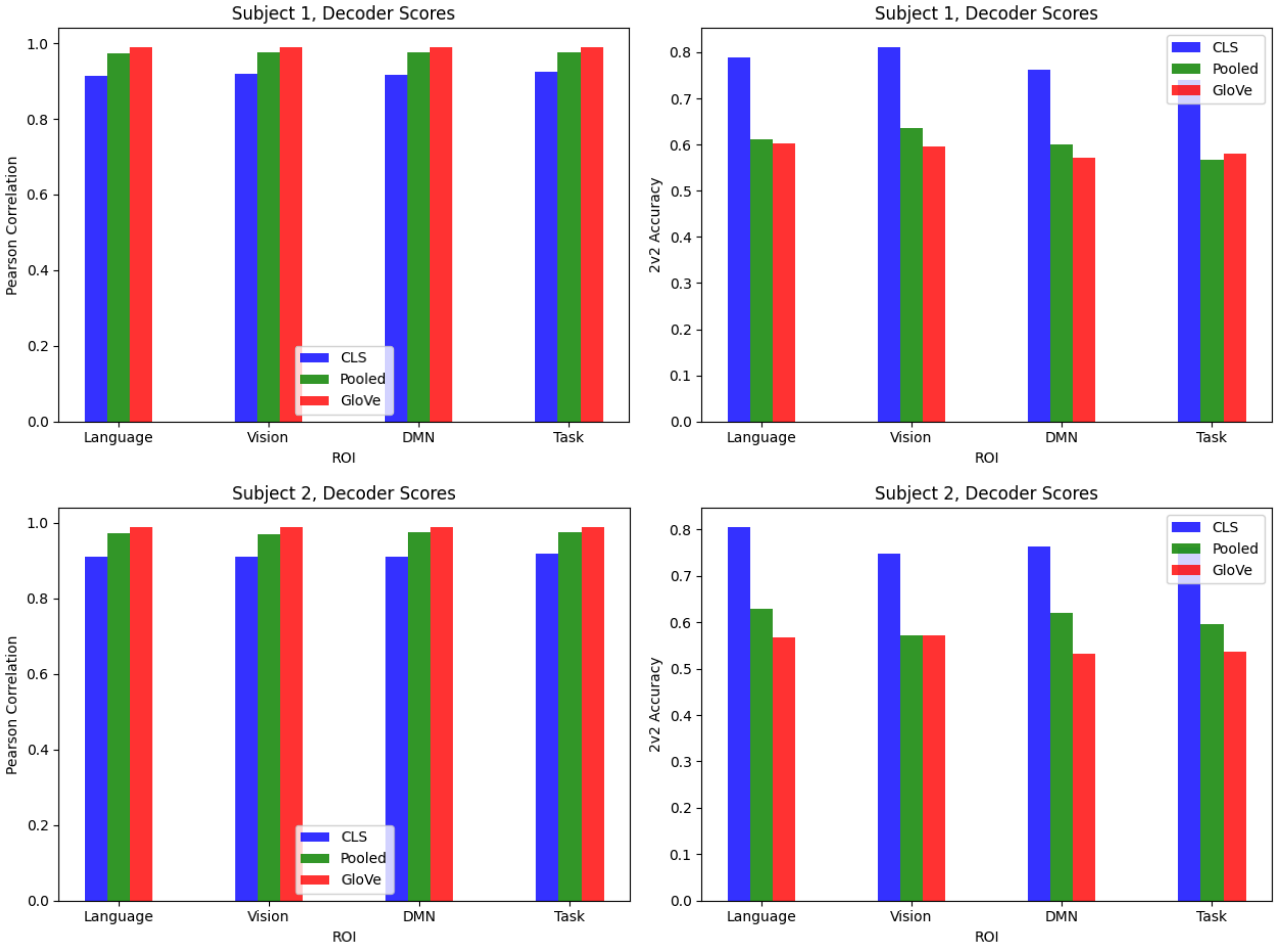
However, the choice of word embeddings may also depend on the specific architecture of the brain encoding and decoding model. For example, if the model uses a convolutional neural network (CNN) architecture, then GloVe embeddings may be more effective because they are trained to capture local semantic relationships between words. Alternatively, if the model uses a

recurrent neural network (RNN) architecture, then CLS or Pooled embeddings may be more effective because they are specifically designed to capture the overall meaning or sentiment of a sentence.

Therefore, the choice of word embeddings for a brain encoding and decoding model using fMRI data with sentence stimuli depends on the specific research question, task, and architecture of the model. It may be beneficial to compare the performance of different word embeddings in the model to determine which one yields the best results.

Comparing the performance of Decoder:

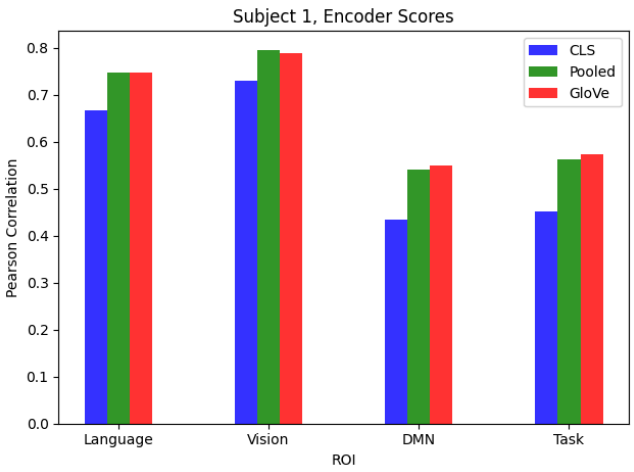
Given Below are the Pearson Correlation and 2v2 Accuracy Scores for the Ridge Regression model trained on the sentence inputs embedded by BERT CLS, Pooled embeddings, and GloVe embeddings for both the subjects.

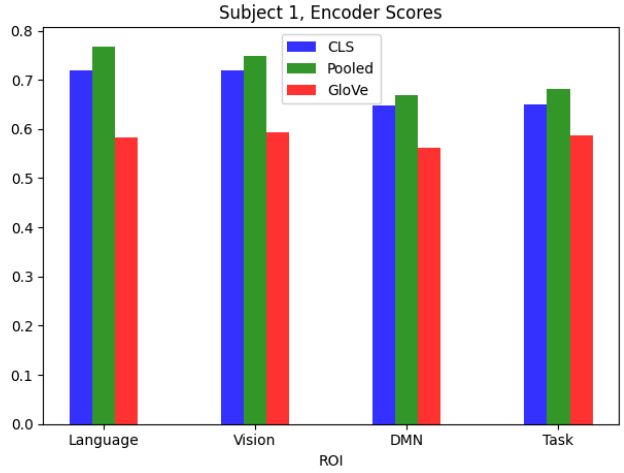


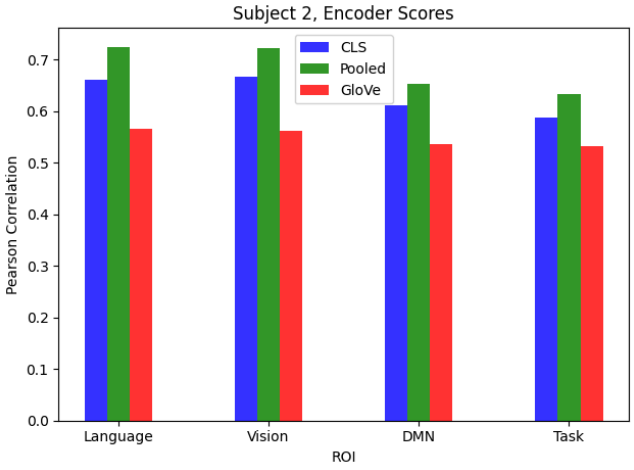
We can see from the graphs above that the predicted sentence embeddings have decent correlations when compared to the test data. When compared using the 2v2 Metric, the CLS embeddings show a very high score, indicating that the decoding model performed best when the sentence was embedded using CLS embeddings. In general, CLS (and Pooled as well, to a lesser extent) embeddings are expected to perform better when deployed in a decoding model as CLS embeddings capture the context of the whole sentence. The opposite is expected of CLS embeddings when used for an encoder. This is because while creating word embeddings, the modifications made cause it to lose information from the original sentence.

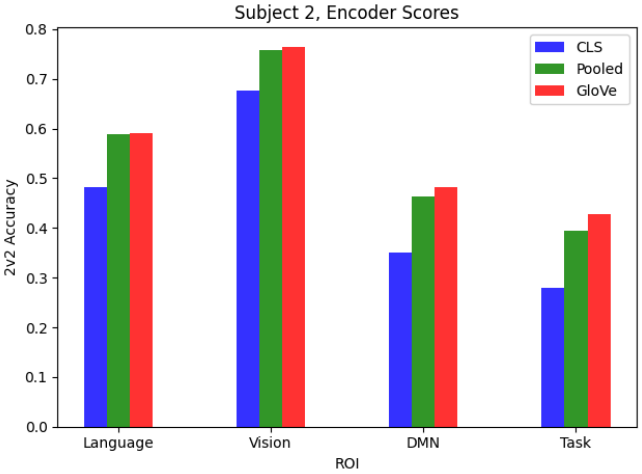
Comparing the performance of Encoder:

Given Below are the Pearson Correlation and 2v2 Accuracy Scores for the Ridge Regression model trained on the sentence inputs embedded by BERT CLS, Pooled embeddings, and GloVe embeddings for both the subjects.









The results of the performances of the three embeddings are not entirely in line with the expected values we set above. However, we do see the performance of the GloVe embeddings getting better in relation to the other two when compared to the relative performances in the decoder. The slight deviations from the expected results may be due various factors such as the limited corpus the model has been trained on. (i.e. inadequate training data).

Hence, we can conclude that for building an encoding model, it is a better idea to go with GloVe embeddings, and CLS embeddings are a better option when we have to build a decoding model.