

NLP HW2- Report

Authors:

Daniel Shemesh (211388251)

Eylon Efraim (207427428)

The goal of this project was to build a model which can recognize entities in text. We used two models to accomplish this task: a K-Nearest-Neighbors model and a Feed-Forward Neural Network.

Training:

K-Nearest-Neighbors (KNN) Model:

We used the KNN model to classify text entities. The KNN model is a supervised learning algorithm which takes a dataset of features and labels and predicts labels for new data based on its similarity to the training data. We calculated the vector representation of words and their contexts using Word2Vec, a pretrained model for converting words to vectors. We then used the KNN model to classify the entities based on the vector representations.

Feed-Forward Neural Network (FFNN):

We used a Feed-Forward Neural Network to classify text entities. The FFNN is a supervised learning algorithm which takes a dataset of features and labels and predicts labels for new data based on the weights of the neurons in the network. We calculated the vector representation of words and their contexts using Word2Vec, a pretrained model for converting words to vectors. We then used the FFNN model to classify the entities based on the vector representations.

Test:

K-Nearest-Neighbors (KNN) Model:

The KNN model had an F1 score of 0.95.

Feed-Forward Neural Network (FFNN):

The FFNN model had an F1 score of 0.95.

Conclusion:

The FFNN model performed better than the KNN model on this task, with an F1 score of 0.95. It is possible that the FFNN model is better suited to this task due to its ability to better generalize from the training data, however there are some improvements that could be made to the Feed Forward model that would increase his performance even more.