Cairo University

Faculty of Computers & Artificial Intelligence

CS462

Natural Language Processing

Assignment 3

Word embedding using Word2Vec

Name: Abdelrahman Ibrahim Ibrahim

ID: 20170139

MLP Classifier (tf-idf)

Using **tf-idf** and MLP classifier I got the following Results.

- Accuracy: 85.5%
- Precision: 0.92%
- Recall: 0.78%

Training Word2Vec model

Parameters that gave best results

- Window: 4 – 5 gave me best results
- Size: size of feature vector affected the results much, values tried are 40,50,60 and 70, best value for size of vector was 128.
- Min_count = 3 – 4 gave best results, default min_count which is 1 gave worse results (but not that much difference)
- Iterations: 20-30 iterations gave good results but not much difference.

MLP Classifier (Word2Vec)

Results

- Different values of alpha gave me some how similar results with small variations. (alpha = [1e-3,1e-5]
- Multiple Hidden Layers sizes also gave similar results.
- **Accuracy**: 81-82%
- **Precision**: 82%
- **Recall**: 82%

SVM Classifier (Word2Vec)

Results:

- Using a linear kernel gave an **Accuracy** of 78%, **Precision** of 81% and **Recall** of 77%.
- Using RBF kernel gave better results, Maxed **Accuracy** 84% (varies), **Precision** of 85% and **Recall** of 83%.
- Also increasing the number of iterations gave little bit better results.

Logistic Regression (Word2Vec)

Results:

- **Accuracy**: 84%
- **Precision**: 85%
- **Recall**: 82%

Summary

TF-IDF gave better results with the MLP classifier (85.5%) accuracy, while using **Word2Vec** for word embedding didn't have as much accuracy as TF-IDF.

Best model with word embedding is either SVM or Logistic Regression as MLP performed poorly with word embedding as its input.