IS 584 Course Term Project Preliminary Results and Benchmarking

Özkan Tuğberk Kartal
Electrical-Electronics Engineering
METU
Ankara, TURKEY
tugberk.kartal@metu.edu.tr

I. Introduction

The purpose of this document is to give explanations regarding Baselines, Preliminary Results, Benchmarking, Github Repository and Weights and Biases (WANDB) Experiment Page of IS 584 Course Term Project.

II. Baselines

As a baseline comparison model, Global Vectors for Word Representation (GloVe) can be used [1]. In order to use GloVe pre-trained word vectors, they should be downloaded from the link [2] and unzipped.

The link of the used dataset is [3]. The texts to be used as inputs in this dataset are titles combined with meta reviews. By using Natural Language Toolkit (NLTK), these input texts are tokenized and English stop words, punctuation marks, None, nan, NA and N\A values are removed from these tokens. The texts to be used as outputs in this dataset are decisions of seven different types which are Accept, Accept (Oral), Accept (Poster), Accept (Spotlight), Accept (Talk), Invite to Workshop Track, Reject.

The classifier trained with GloVe embeddings is a Multi-layer Perceptron (MLP) classifier. MLP classifier has a hidden layer of size 128. Training and testing ratios are 0.8 and 0.2, respectively. Training is done for maximum 50 epochs with early stopping patience being equal to 3. In every epoch, Classifier Checkpoints, Accuracy, F1-Score, Area Under the Receiver Operating Characteristic (ROC) Curve and Confusion Matrix are calculated and stored to evaluate the performance of the MLP classifier.

After the training is done; the final values for Accuracy, F1-Score, and Area Under the Receiver Operating Characteristic (ROC) Curve are equal to 0.62218, 0.58854, and 0.73483, respectively, and the Confusion Matrix is given by Figure 1. Much detailed results can be found in [4] and [5]. Importantly, the related .ipynb file can be seen in [6].

III. Preliminary Results

The main model architecture used is a Capsule Neural Network architecture in PyTorch framework [7], [8].

The link of the used dataset is [3]. The texts to be used as inputs in this dataset are titles combined with meta reviews. By using Natural Language Toolkit (NLTK), these input texts are tokenized and English stop words, punctuation marks, None, nan, NA and N\A values are removed from these tokens. The texts to be used as outputs in this dataset are decisions of seven different types which are Accept, Accept (Oral), Accept (Poster), Accept (Spotlight), Accept (Talk), Invite to Workshop Track, Reject.

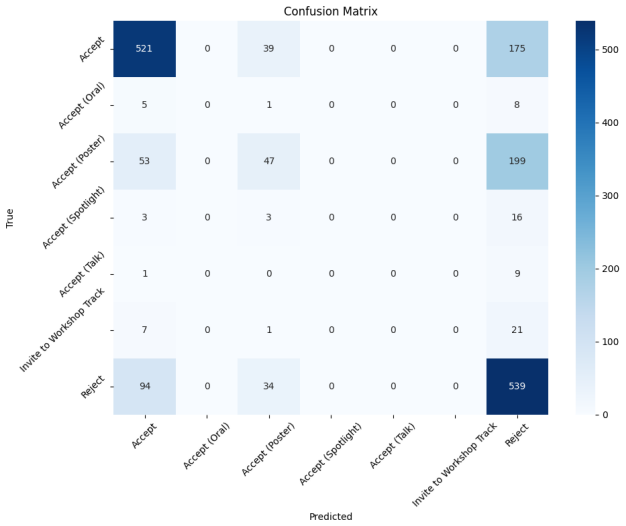

Fig. 1. GloVe Confusion Matrix

Capsule Layer Architecture has the following parameters

- num capsules: Number of output capsules in the layer = 2
- dim capsule: Number of dimensions in each output capsule's vector = 16
- input dim: Number of dimensions in each input capsule's vector = 3
- input len: Number of input capsules = 4
- routing iters: Number of routing iterations = 3

Capsule Layer is preceded with an embedding layer and two convolutional layers and is followed by a linear output layer. The embedding layer is of type vocabulary size x embedding dimension (20002 x 32), the first convolutional layer is of type embedding dimension x 64 (32 x 64), the second convolutional layer serving as primary capsules is of type 64 x (input dim x input len) (64 x 12), the capsule layer serving as dense capsules is of type 1 x 24 x 2 x 16 x 3 and the linear output

layer is of type 512 x output length (512 x 7).

Training and testing ratios are 0.8 and 0.2, respectively. Training is done for 50 epochs with learning rate = 1e-3 and batch size = 128. In every epoch, Model Checkpoints, Cross Entropy Loss, Accuracy, F1-Score, Area Under the Receiver Operating Characteristic (ROC) Curve and Confusion Matrix are calculated and stored to evaluate the performance of the Capsule Neural Network architecture.

After the training is done; the final values for Cross Entropy Loss, Accuracy, F1-Score, and Area Under the Receiver Operating Characteristic (ROC) Curve are equal to 0.37507, 0.54561, 0.53338, and 0.71804, respectively, and the Confusion Matrix is given by Figure 2. Much detailed results can be found in [4] and [5]. Importantly, the related .ipynb file can be seen in [9].

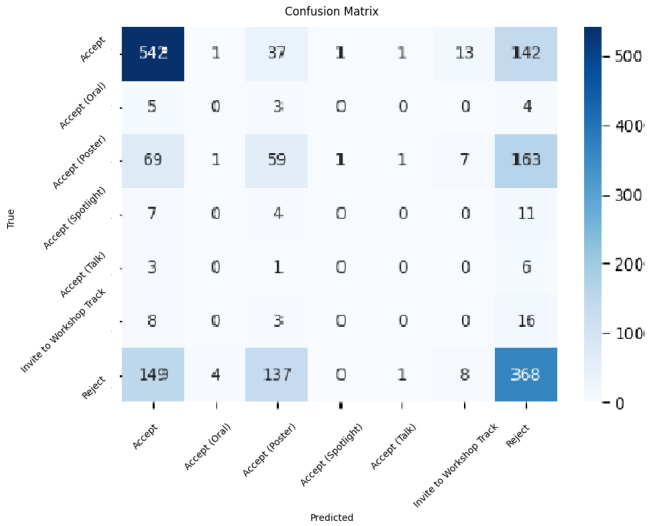

Fig. 2. Capsule Neural Network Confusion Matrix

IV. Benchmarking

Comparing the results of GloVe and Capsule Neural Network in terms of Accuracy, F1-Score, Area Under the Receiver Operating Characteristic (ROC) Curve and Confusion Matrix, it could be said that similar performances are reached. To improve the performance of Capsule Neural Network, LSTM, Attention and/or Transformers based additions can be implemented.

V. GITHUB

The Github link of the repository is given by [4].

VI. WANDB

The link of the WANDB experiment page is given by [5].

References

[1] 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

[2] J. Pennington, R. Socher, and C. D. Manning, "GloVepre-trained word vectors," 2014. [Online]. Available: http://nlp.stanford.edu/data/glove.6B.zip

[3] "IS 584 Course Term Project Dataset," 2025. Available: https://drive.google.com/file/d/1nJdljy468roUcKLbVwWUhMs7teirah75/view

[4] Ö. T. Kartal, IS584CourseTermProject. Online. Available: https://github.com/oztuka/IS584CourseTermProject/

[5] Ö. T. Kartal, IS 584 Course Term Project. Online. Available: https://wandb.ai/eabu2ss-metu-middle-east-technical-university/IS%20584%20Course%20Term%20Project

[6] Ö. T. Kartal, IS584CourseTermProject. Online. Available: https://github.com/oztuka/IS584CourseTermProject/blob/main/notebooks/Baselines.ipynb

[7] Wikipedia contributors, Capsule neural network — Wikipedia, The Free Encyclopedia. 2024. [Online]. Available: https://en.wikipedia.org/wiki/Capsule_neural_network

[8] S. Sabour, N. Frosst, and G. E. Hinton, Dynamic Routing Between Capsules. 2017. [Online]. Available: https://arxiv.org/abs/1710.09829

[9] Ö. T. Kartal, IS584CourseTermProject. Online. Available: https://github.com/oztuka/IS584CourseTermProject/blob/main/notebooks/Preliminary_Results.ipynb