

# Assignment 4 Report

M.M.Yaseen

April 2024

## 1 ELMO

- Two models, one for forward language modeling, and other for backward language modeling were implemented using LSTM.
- each of those models were trained for 10 epochs using Adam Optimizer with learning rate of 0.001 and a cross entropy loss function.
- No pretrained embeddings like Word2vec or Glove were used rather an embedding layer was used to generate embeddings to train these models.

## 2 DownStreamClassification

- using pretrained embeddings, Downstream Classification was performed, by passing it through forward and backward pretrained models, the concatenate them to obtain the embedding, this was done for each lstm layer and embedding layer yielding three variables.
- All three models were test below are the results and metrics obtained using the three models, for model using scalars Adam optimizer was used with learning rate of 0.001.
- for trainable function all 3 embeddings  $e_1, e_2, e_3$ , were concatenated and fed to a Linear Layer which finally results in an embedding. Adam Optimizer with learning rate of 0.0001 was used.

### 3 Results

```
(virtual) [yaseen@legion a4]$ python evaluate.py
```

```
100%|
trainable lambdas
f1 score :0.8340005843110445
recall score: 0.8340789473684211
accuracy score: 0.8340789473684211
confusion matrix:
[[1630    86    71   113]
 [ 100 1692    30    78]
 [ 128   40 1405   327]
 [   95   47  146 1612]]
```

```
100%|
freezed lambdas
f1 score :0.8360458972009114
recall score: 0.8364473684210527
accuracy score: 0.8364473684210526
confusion matrix:
[[1543  114  134  109]
 [  52 1776   33   39]
 [  83   52 1506  259]
 [  84   69  215 1532]]
```

```
100%|
trainable function of embeddings
f1 score :0.8488309941192136
recall score: 0.8488157894736841
accuracy score: 0.8488157894736842
confusion matrix:
[[1691    52    80    77]
 [  74 1715    47    64]
 [ 122   32 1480   266]
 [ 118   38  179 1565]]
```

## 4 Analysis

- out of all the trainable function yielded a good accuracy compared to others.
- this outperforms SVD, and Word2vec with small context windows, where as give similar performance to Word2vec with context window fo 4 or 5.
- However, these embeddings can be further trained compared to word2vec using trainalbe function to yeild a better accuracy across various tasks .