Natural Language Processing

Assignment 2

Q1) Sentiment analysis using deep learning

This question uses following dataset of Urdu sentiment analysis. The class labels are P(positive) and N (Negative)
https://github.com/MuhammadYaseenKhan/Urdu-Sentiment-Corpus/blob/master/urdu-sentiment-corpus-v1.tsv

Implement following sequence based deep learning models for the same task of sentiment analysis. Perform binary text classification.

RNN
GRU
LSTM
BiLSTM

You can implement these models in Keras or Pytorch. Split the data into train and test set. Use 75% for training and 25% for testing.

For each of these models, try following hyper parameters and report the best results with parameter values.
Number of layers = 2 or 3.
Dropout rate, 0.3 or 0.7
So you will have 2 *2 = 4 different sets of parameters.

Calculate accuracy, Precision, Recall and F-score for all classifiers and report the results in table.

Also report parameter values which were used to get the results.

Q2) Sentiment Analysis using word embedding

This question uses same dataset of Q 1. Perform the task of **binary classification** on the dataset. Choose one classifier from deep learning models implemented in Question 1 based on best results on F-measure for binary classification.

Use following embedding for vector representation and report the results. You **need to train** the embeddings yourself on the given Urdu dataset.

1) WordToVec

https://mccormickml.com/2016/04/12/googles-pretrained-word2vec-model-in-python/

2) Glove
   https://nlp.stanford.edu/projects/glove/
   https://medium.com/analytics-vidhya/basics-of-using-pre-trained-glove-vectors-in-python-d38905f356db
3) Fasttext
   https://blogs.sap.com/2019/07/03/glove-and-fasttext-two-popular-word-vector-models-in-nlp/
   https://fasttext.cc/docs/en/english-vectors.html
4) Elmo (it creates embeddings for sentences, so use entire tweet as imput to get the vector) https://github.com/HIT-SCIR/ELMoForManyLangs

Calculate accuracy, Precision, Recall and F-score for all classifiers and report the results in tables. For example, if LSTM (use any fixed hyper-parameters that gave best results) had best overall results among deep learning models in your assignment 1 then make the following table of results.

	LSTM (without embeddings)	LSTM with WordToVec	LSTM with Glove	LSTM with Fasttext	LSTM with Elmo
F-score					
Accuracy					
Precision					
Recall					

Submission

Submit the code files and result tables as zip file on Google classroom.