IT412: Natural Language Processing

Assignment 8: Hate Speech Classification

Instructor: Prasenjit Majumder

Learning Outcome: At the end of this assignment you will learn to perform hate speech classification using deep learning.

1 Problem description

Text classification is the task of assigning a text to its proper category. The two main techniques that are used for text classification are supervised and unsupervised. In supervised learning we use labelled data to train our classifier, and then testing is performed. In unsupervised learning text clustering is performed. Text clustering is the task of grouping a set of unlabeled texts in such a way that texts in the same group (called a cluster) are more similar to each other than to those in other clusters.

2 Implementation

2.1 Dataset

- For this assignment we will use Jigsaw Toxic comment classification dataset.
  https://drive.google.com/file/d/18l6IwSqavnqtLQpVnrRqOugZf9XkhEAN/view?usp=sharing
- It consists of 5 categories:
  1. toxic
  2. severe_toxic
  3. obscene
  4. threat
  5. insult
  6. identity_hate
- The description of the files is given below:
  1. train.csv - the training set, contains comments with their binary labels
  2. test.csv - the test set, you must predict the toxicity probabilities for these comments. To deter hand labeling, the test set contains some comments which are not included in scoring.
  3. sample_submission.csv - a sample submission file in the correct format
  4. test_labels.csv - labels for the test data; value of -1 indicates it was not used for scoring;

2.2 Exercise

- Use BiLSTM to get the context vector
- Initialize the word embedding layer using Normal Distribution (Mean 0 and Variance 1)
- Pass the Context vector through Full connected layer and apply softmax to get the probabilities
- Use Cross Entropy as loss function
- Report the Macro F1 score for your model

3 References

- https://www.kaggle.com/jhoward/improved-lstm-baseline-glove-dropout
- https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html
- https://pytorch.org/docs/stable/generated/torch.nn.Embedding.html
- https://www.kaggle.com/jhoward/improved-lstm-baseline-glove-dropout

4 Submission

- You have to submit your assignment in Jupyter notebook with proper comments and explanation of your approach.
- The submission deadline for this assignment in **22nd November 2021 at 11 PM**