CS772: DL4NLP Assignment-3 Report

Team Id : 25
Debabrata Biswal, 203050024
Ashish Aggarwal, 203050015
Keshav Agarwal, 203050039

Date: 12/04/2020

1. Problem Definition:

- Identifying the rating of a review using sentiment analysis
- Input dataset contains reviews and their rating score between 1 to 5 - low to high.
- Given a review, we predict the rating score. Examples:
    - This product is really awesome should have a rating of 5.
    - This product is pathetic should have a rating of 1.

2. Architecture:

1. **Input Preprocessing Layer**: We are doing the following operations:
    a. Convert all characters in input string to lowercase.
    b. Remove all punctuations.
    c. Tokenize the given input strings using NLTK library word_tokenize function.
    d. Remove all stopwords excluding negative words using the NLTK library stopwords function.
    e. Encode the vocabulary word to numbers.
    f. Split the train data to train and validation set of 80-20% using scikit learn library train_test_split function.

2. **Embedding Layer**: We are comparing two setting of configuration.
    a. Pre-Trained Embedding: We are using pre-trained FastText embedding of length 300 for initializing the weights of embedding layer and then training on our input set.
    b. No Pre-Trained Embedding: We are directly training the weights of embedding layer on input set using gensim library word2vec function.
3. NN Layer: We are comparing results of 5 different neural network architectures.
    a. RNN
    b. LSTM
    c. Bi-LSTM
    d. GRU
    e. Bi-GRU
4. Dense Layer: We are using this layer to reduce the dimension of output to 5 classes.
5. Softmax Layer: To get the probability distribution over set of 5 classes.

output: [(None,5)]

Fig: Architecture Diagram

3. Prediction and Evaluation:

- Four evaluation metrices captured for comparing all 5 NN architecture performance:
  - o Precision
  - o Recall
  - o F1-score
  - o Accuracy
- We are using scikit-learn library classification_report , confusion_matrix and precision_recall_fscore_support functions to capture the performance metrices.

4. Performance Comparison:

a. Pre-Trained Embedding:

Model	Dense_layers	Precision	Recall	F1_score	Accuracy
Bi-GRU	0	0.33	0.58	0.42	0.58
Bi-GRU	1	0.69	0.62	0.64	0.62
Bi-GRU	2	0.33	0.58	0.42	0.58
Bi-LSTM	0	0.33	0.58	0.42	0.58
Bi-LSTM	1	0.68	0.57	0.61	0.57
Bi-LSTM	2	0.33	0.58	0.42	0.58
GRU	0	0.33	0.58	0.42	0.58
GRU	1	0.33	0.58	0.42	0.58
GRU	2	0.33	0.58	0.42	0.58
LSTM	0	0.33	0.58	0.42	0.58
LSTM	1	0.66	0.52	0.57	0.52
LSTM	2	0.33	0.58	0.42	0.58
RNN	0	0.33	0.58	0.42	0.58
RNN	1	0.67	0.57	0.61	0.57

Model					
RNN	2	0.33	0.58	0.42	0.58

b. Without Pre-Trained Embedding:

Model	Dense_layers	Precision	Recall	F1_score	Accuracy
Bi-GRU	0	0. 69	0.62	0.65	0.62
Bi-GRU	1	0.69	0.63	0.65	0.63
Bi-GRU	2	0. 69	0.62	0.65	0.62
Bi-LSTM	0	0.70	0.61	0.64	0.61
Bi-LSTM	1	0. 69	0.62	0.64	0.62
Bi-LSTM	2	0.68	0.61	0.63	0.61
GRU	0	0.33	0.58	0.42	0.58
GRU	1	0.33	0.58	0.42	0.58
GRU	2	0.33	0.58	0.42	0.58
LSTM	0	0.68	0.61	0.64	0.61
LSTM	1	0.69	0.56	0.60	0.56
LSTM	2	0.68	0.54	0.59	0.54
RNN	0	0.67	0.56	0.60	0.56
RNN	1	0.66	0.58	0.61	0.58
RNN	2	0.67	0.64	0.65	0.64

5. Qualitative Analysis:
a) Not word Holds a strong negative sentiment:
   - Review 'product is not bad' should have less probability of class 1 compared to 'product is bad'.
   - Review 'product is good' has a correct review of 4 or 5 and reduces to 2 or 3 when not is added.

Review	Rating	Class 1	Class 2	Class 3	Class 4	Class 5
product is not bad	1	0.508	0.354	0.101	0.023	0.015
product is bad	1	0.336	0.238	0.161	0.131	0.134
product is not good	2	0.206	0.277	0.268	0.157	0.093
product is good	4	0.047	0.062	0.143	0.38	0.368

b) There is noise in training data. But our model is predicting right. Example:
  o In the first sentence of the below image user has positive review but ratings given in training set is 1 which is wrong.
  o Likewise in the last sentence user has negative sentiment but rating label is 5. Our model is predicting 1 correctly because of words like expensive, no which has negative sentiment associated with them.

	Unnamed: 0	reviews	ratings	pred
19843	19843	Clean and bright!! Just what I needed!!	1	5
5240	5240	This is the best brand of blush.love the colors too.	1	5
2207	2207	Used it before, no change. IT'S expensive!9	5	1