CS772: DL4NLP
Assignment Evaluation

Team Id : 25
Debabrata Biswal, 203050024
Ashish Aggarwal, 203050015
Keshav Agarwal, 203050039
Date: 12/04/2020

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.

Architecture

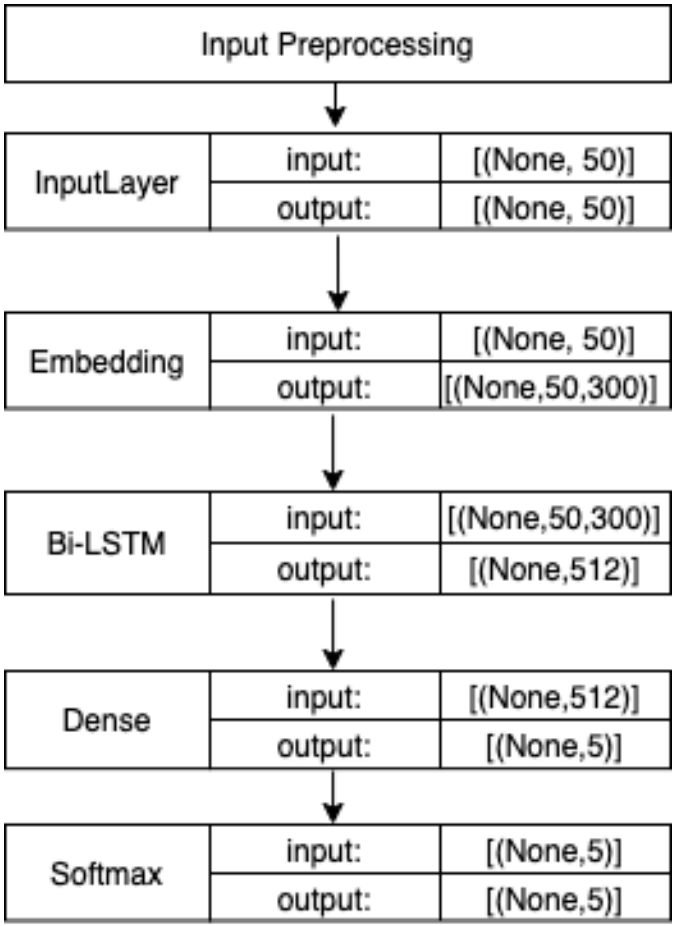

Diagram generated via Draw.io

Libraries

Data preprocessing - scikit-learn for validation split - NLTK library	scikit-learn function - train_test_split NLTK functions - Stopwords - word_tokenize
Word to integer encoding - manually	N.A.
Word embedding - Word2Vec using gensim - FastText	gensim function - Word2Vec
Neural Network - TensorFlow – keras	keras function - EarlyStopping
Prediction and Evaluation - keras - scikit-learn	scikit-learn function - classification_report - confusion_matrix - precision_recall_fscore_support

Table 1: Model Performance Comparison

Model	With Pre-Trained Embedding	Without Pre-Trained Embedding
RNN	0.67 0.57 0.61 0.57	0.66 0.58 0.61 0.58
LSTM	0.66 0.52 0.57 0.52	0.69 0.56 0.60 0.56
Bi-LSTM	0.68 0.57 0.61 0.57	0.69 0.62 0.64 0.62
GRU	0.33 0.58 0.42 0.58	0.33 0.58 0.42 0.58
Bi-GRU	0.69 0.62 0.64 0.62	0.69 0.63 0.65 0.63

<Precision> <Recall> <F1-Score> <Accuracy>

Confusion matrix for the best model

Classified

Actual	1	2	3	4	5
1	748	331	132	20	40
2	153	245	167	38	27
3	60	225	424	149	53
4	26	101	382	497	399
5	113	123	361	934	4253

Qualitative analysis with some examples

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

- 'not' word holds a strong negative sentiment value
- 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.