Abstract

In this paper, we describe an approach to perform sentiment analysis on tweets from users from major airlines in US in 2015. The aim of the project was to investigate whether the cooperation of Bidirectional Long Short-Term Memory (Bi-LSTM) and attention will provide high performing metrics on predicting the sentiments of the tweets. In order to provide a better context for rating the performance of the proposed model, a simple vector machine model, known to be a very efficient machine learning model was implemented as a base model. The results from the experiment indicated an accuracy score of 77.1% for the Bi-LSTM and attention model as opposed to a 56% accuracy score for the SVM.

1.Introduction

Sentiment analysis (SA) is the key to promote the advancement of artificial intelligence and is applied by plenty of corporates to achieve their business mission. In addition, Twitter data is a good source since people prefer to write down feelings and comments on microblogs. Therefore, this paper focuses on sentence-level sentiment analysis of Twitter data. Sentence-level sentiment classification supposes that every sentence expresses a emotion and aims to decide the overall sentiment about different airlines.

This paper proposes a deep learning neural network including a Bidirectional Long Short-Term Memory and a sentence-level Attention. Based on the goal of this paper, the research question is raised as:

> *Could the cooperation of Bi-LSTM and Attention provide a high accuracy of predicting sentiments in the comments from Twitter US Airline Data?*

Contribution

This paper proposes an efficient neural network for sentiment analysis by combining attention to Bi-LSTM. This cooperation contributes to a better sentence representation with paying attention to every word in comments and successfully improve the accuracy of predicting.

2.Related Work

2.1 Hierarchical Long Short-Term Memory

Hierarchical Long Short-Term Memory (HLSTM)is applied to extract the long-range dependency during the process of retweeting by Huang, Cao and Dong (2016). Two layers are in the HLSTM: one word-level layer works for representation of every single tweet and the other tweet-level investigates the long-range context of every corresponding tweet. Additionally, context features captured from original tweets.

2.2 Hierarchical Long Short-Term Memory with Attention

Chen, Sun, Tu, Lin and Liu (2016) implement HLSTM to produce the representation of sentence and document, and a layer of attention works with information of user and product to improve a document-level sentiment analysis.

2.3 Convolutional Neural Network with Long Short-Term Memory

Zhou, Sun, Liu and Lau (2015) integrate two neural structures as inputing the output of CNN to LSTM: CNN generates phrase representation and LSTM produces sentence representation.

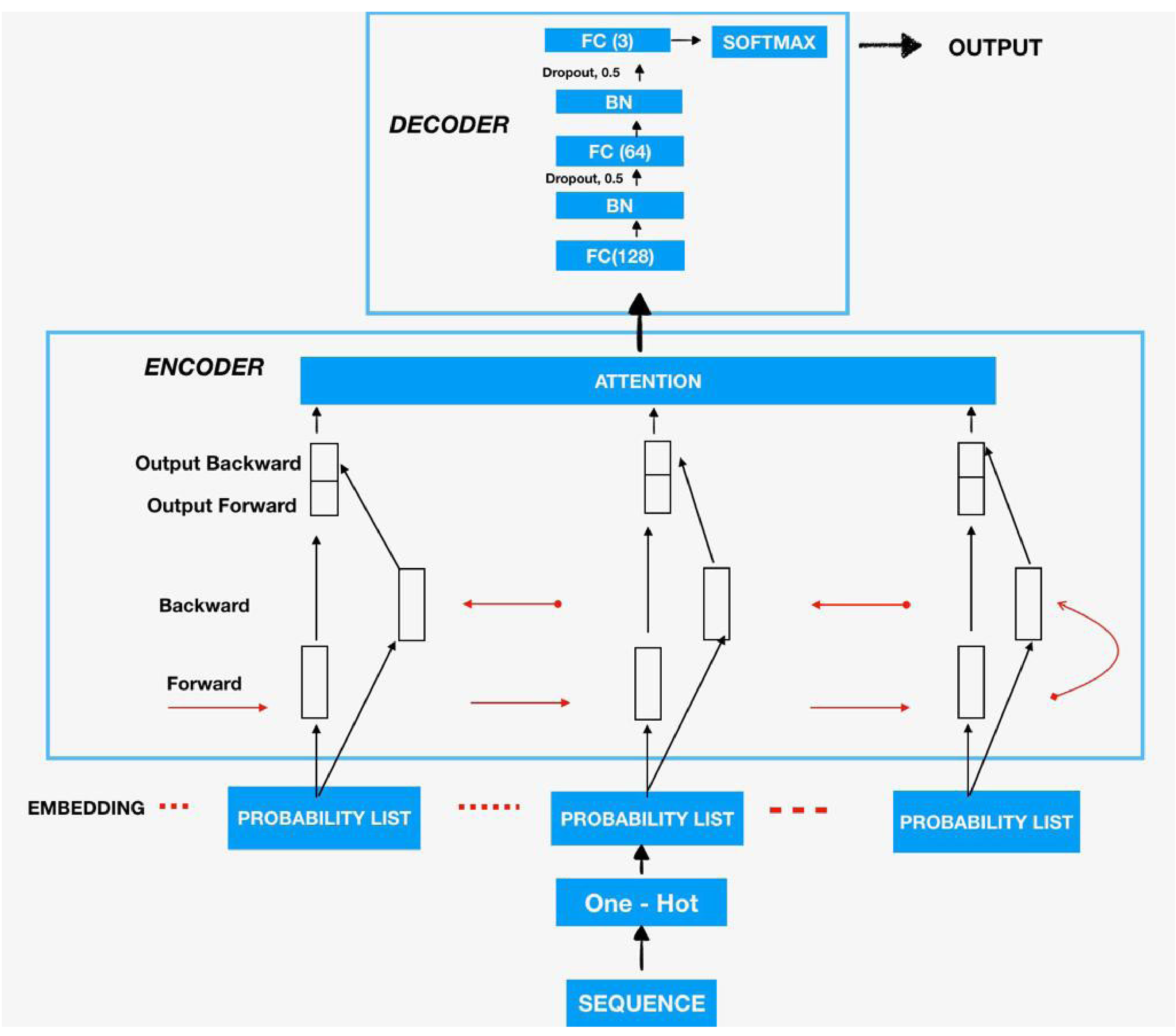
Difference

Compared to the one implementing HLSTM, this paper apply a different Bi-LSTM and adds one more layer of attention. This structure may be more comprehensive when capturing useful features as much as possible. However, additional context features obtained from parent tweets are not included in this project due to the limitation of data which has no information of retweeting.

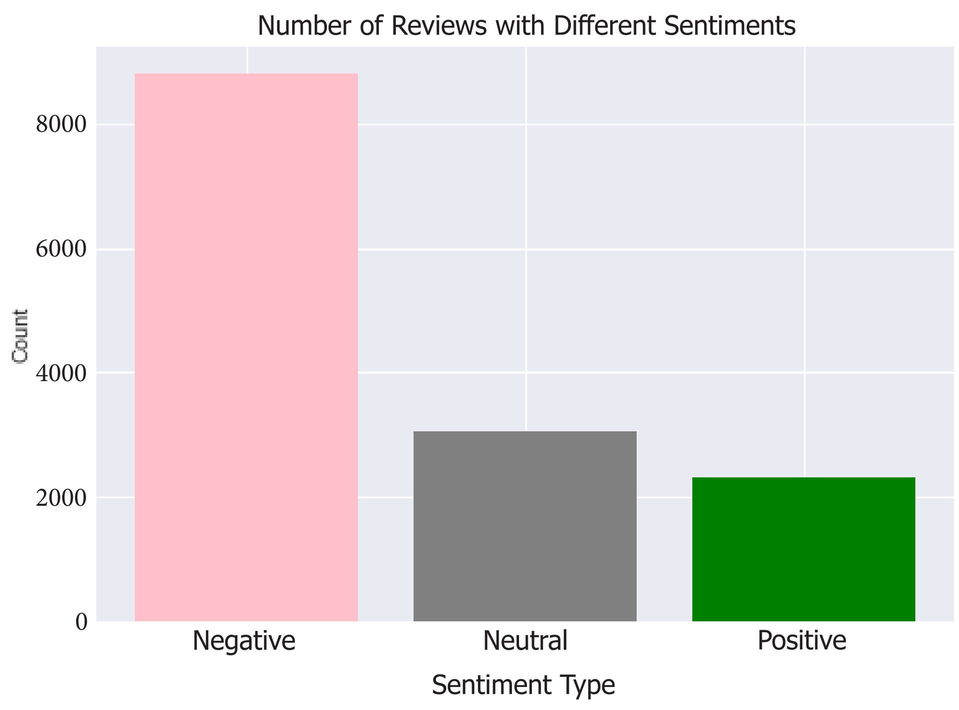
In case of the one with HLSTM and user product attention, similarly this paper chooses to use Bi-LSTM to fully use information from past and future. The reason for not implementing HLSTM is that the specific topics of sentiment analysis are different. This paper focuses on the sentence level and discovers the emotional polarities of each short-length tweets rather than the document-level sentiment analysis.

3. Method

Model



Data

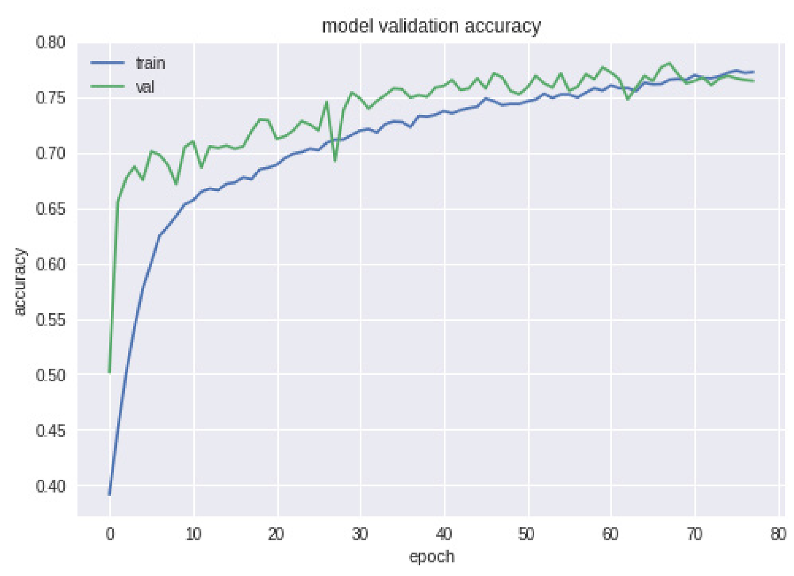


4.Experiment

We utilized the Colab and its GPU as the environment of Python. Regarding the packages that we imported, we use Keras to develop our neural network, which includes pretrained word embedding(glove), Bi-LSTM (256 neurons, dropout=0.5, Recurrent Dropout=0.25), Attention, two Fully Connected Layer (128 neurons and 64 neurons, dropout=0.5) and Batch Normalisation based on the grid search for the hyperparameter tuning.

Furthermore, the loss function for our model is categorical_crossentropy (Diederik P. Kingma, Jimmy Ba, 2016) and the optimizer is the Adam (Kingma and Ba, 2014). For the experiment, we run the model in maximum epoch of 300; however, we also add early stopping function of patience=10. Finally, in the testing set, we try to use confusing matrix to explain the result and the future improvement of this model.

5.Result and Discussion



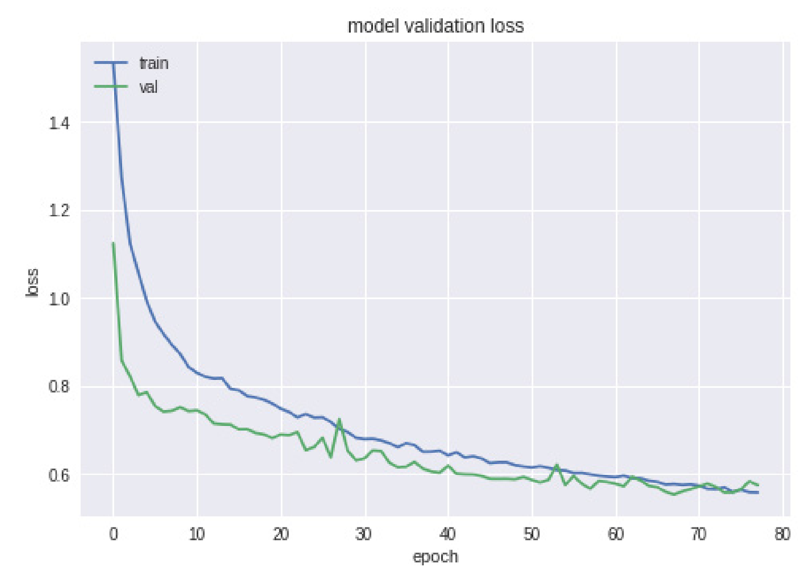


Table 1: Results of Proposed Model in Test Set

(a) Confusing Matrix of Bi-LSTM + Attention

	Negative (Predicted)	Neutral (Predicted)	Positive (Predicted)
Negative (Real)	832	61	25
Neutral (Real)	115	170	25
Positive (Real)	75	33	128
Total number of data = 1464			

(b) Metrics of Bi-LSTM + Attention

	Precision	Recall	F1 Score
Negative	0.81	0.90	0.85
Neutral	0.64	0.54	0.58
Positive	0.71	0.54	0.61
Overall accuracy = 0.771			

Table 2: Result of Baseline in Test Set

(a) Confusing Matrix of SVM

	Negative (Predicted)	Neutral (Predicted)	Positive (Predicted)
Negative (Real)	789	114	18
Neutral (Real)	279	18	13
Positive (Real)	207	20	39
Total number of data = 1464			

(b) Metrics of SVM

	Precision	Recall	F1 Score
Negative	0.62	0.72	0.86
Neutral	0.12	0.08	0.06
Positive	0.23	0.07	0.04
Overall accuracy = 0.56			

Comparison

Compared to SVM, the neural network(NN) produces moderate higher results in overall accuracy, recall, precision and F1 score in all types of labels. The reason is that Bi-LSTM and Attention layers can capture the significant word or vectors that represent the input words. That is to say, after NN get the significant representation of sequences, it can easily predict the words. According to Sboev et.al (2018), NN can basically produce a higher accuracy in a complex situation than traditional machine learning model. In the small dataset, traditional machine learning algorithm could work better than NN. However, in the large dataset and a complex situation, for instance a NLP task, it is hard for traditional machine learning to produce a good result.

Possible Reason1: Regarding the result of NN and SVM model, it can be understood that since the data count for each label is not equal. For instance, in the training set, the negative data is 7434; the neutral data is 2510 and the positive data is 1914.

Possible Reason2: In a specific scenario, humans will use some specific words or sentences meaning some specific situations. For instance, the comment : "@VirginAmerica when can I book my flight to Hawaii????".

6.Conclusion & Future Work

- Results show Bi-LSTM and Attention gives a better performance for the prediction of sentiments of tweets than SVM
- Bi- LSTM was initially presumed to be a better predictor since it is able to capture long term tendencies in sentences and compute a better representation of its words from multiple abstract levels.
- SVM on the other hand, captures sparse and discrete features, making it difficult to detect the characteristics of sentences that are key factors in determining the overall sentiment polarity.

Future Works:

Interesting to include other tweet features to determine the influence they have on the models.