© APR 2018 | IRE Journals | Volume 1 Issue 10 | ISSN: 2456-8880

Sentiment Analysis on Movie Reviews using Recurrent Neural Network

SUMESH KUAMR NAIR [1], RAVINDRA SONI[2]

[1,2] *Department of Computer Science and Engineering, Poornima College of Engineering*

Abstract -- In this paper i have done sentiment analysis on IMDB dataset using Recurrent Neural network. Sentiment analysis based on text mining or opinion mining based on different dataset. Sentiment classification is done in three categories- Positive, Negative and Neutral. Text classification is done on the dataset and data preprocessing is done to remove hastags, synonms, acronyms etc. LSTM Recurrent Neural Networks to other algorithms for classifying the sentiment of movie reviews. Recurrent neural network provides high accuracy and polarity as compared to different machine learning classifiers. To address this task deep learning has become popular method. LSTM (Long short-term memory) model has been used which is a modified version of RNN (Recurrent Neural Networks). Recurrent Neural Networks has ability handle sequential data very effectively and without performing any feature engineering it can learn directly from low-level features. Instead of exploring LSTMs abilities and capabilities, main focus was to learn how embedding can help us to understand user expectations from text. Proper pre- processing for data has been implemented. Informal language, contextualization, bad grammatical structure, misspellings are additional complicating factors. Reviews are analyzed as binary classification task, after processing reviews are classified as either negative or positive. Features for training and testing the deep learning model were retrieved by using new method called 'word-vector'. Moreover, effect of sentence length has also been investigated. Sentiment analysis for short sentences becomes difficult because of lack of contextual information. Multiple hidden layers have been used in the architecture. Dropout, Normalization and Parametric Rectified Linear Unit (PReLU) technology has been used to generalize and improve the accuracy of model. Also, the impact of various hyper-parameter has analysed. Different neural network configurations are evaluated. The performance of model is discussed with respect to the input data and model configuration.

Index Terms- Sentiment analysis;PReLU;LSTM;RNN

I. INTRODUCTION

Today extensive datasets are accessible on-line, holding text data or numerical. It has been the major focus for many practitioners and researchers to apply reasonable approaches and techniques and extract useful information from those datasets. Wide range of techniques have been proposed and tested to retrieve information during this time. In addition to text mining and data mining, lately interest for non-topical text analysis have increased drastically and sentiment analysis is part of them.

Sentiment analysis is a process of analyzing the given text in order to find out the emotions in it. Sentiment analysis is about "Text analysis, Information Extraction and Natural Language Processing are kind of tasks which aim towards getting the writer's feelings expressed in negative or positive comments by analyzing sentences or documents" defined by Subhabrata Mukherjee. In simple words, opinion mining is a process of detecting the sentiment of the writer concerning a particular topic. It is a blend of techniques and strategies about distinguishing and detecting subjective information from a text such as opinions and attitudes. Usually, it has been about opinion polarity to find out whether someone has negative, positive or neutral opinion about something.

Internet is a huge source of information for every individual. For instance, investors want to be updated with financial news particularly associated with their investment. Organizations are looking for the news about competitors, suppliers and customer's feedback. This is a cyclic chain, likewise customers have interest in reviews of other customers about the products they are looking for. Researcher made an excessive amount of effort to identify the impacts of this technique on customer insight, trend and financial world. Numerous applications of sentiment analysis came up in multiple domains like sentiment analysis of products reviews, financial news and healthcare. It additionally offers them a superior picture of how they stack up against their competitors.

Technically, sentiment analysis is a unique blend of artificial intelligence and machine learning, allowing organizations to use advanced tools to choose useful and reasonable moves that attract consumers toward their services and products. In order to retain customers, competitors have to track and monitor the interest of customers. Especially, not towards their own products and brands only but also towards their competitors.

Machine learning have seen rapid change in previous two years with significant breakthroughs in deep-learning approaches. Deep neural networks enlivened by the human brain architecture and with enough processing power these models have been shown unbelievable results on many complex problems including Natural Language Processing tasks, even without having excessive domain knowledge. Out there many neural networks are available with their classic abilities like Deep Belief Networks (DBN) with fast inferencing of the model parameters, Convolutional neural networks (CNN), and Recurrent neutral network (RNN). In this work, I will work with LSTM (Long Short- Term Memory). LSTM networks are a type of RNN that uses special units in addition to standard units.

II. RELATED RESEARCH

There has been considerable research done and still going on sentiment analysis subject. Sentiment analysis became part of research at the beginning of 20th century and in 1990, text subjectivity analysis was performed by computational linguistic community [1]. Most of the sentiment analysers used to work by choosing one or hybrid of following approaches.

- Using Vocabulary:

Worked by choosing the important keywords (usually adjectives and verbs) along with modifiers. For example, negative words.

- Machine Learning approach:

Treat sentiment analysis as classification problem, extract features. Train model to determine sentiment.

- Using Rules:

Look for presence of specific words in sentence and define rules based on those words and categorize sentences.

IIn following sections, old approaches for classification of sentiment analysis and difficulties in sentiment analysis has explained.

2.1 Sentiment Analysis Difficulties

Research demonstrates that the task of sentiment analysis is more tough than conventional topic based classification of text, regardless of the fact that we don't have much classes in SA than classes in topic based classification [10]. In this task, usually classification assigned to the text are generally positive or negative. There can be some different binary classes also or multivalued classification. For example, neutral, negative or positive, yet those classes are not as much as in topic based classification. Topic based classification is a bit easier than sentiment analysis because this can be achieved with the use of keywords this could be a reason. On the other hand, this technique doesn't perform well with sentiment analysis [11].

Classification in sentiment analysis is a subjective method but there could be variations in opinions if there are number of observers to test. Interpreting the state of mind of a subject may differ person to person and if someone has only 140 characters or less to express something then its significantly hard to determine the mood [12].

The study of sentiment classification and subjectivity classification is required to perform sentiment analysis properly. In subjectivity classification, it is identified that whether or not provided text data contains opinionated information or factual information. Similarly, sentiment analysis is a process to classify an opinion into negative or positive. In reality, if we consider a product review then it requires in depth analysis of assigned classification because the manufacturer of product is interested to know the details of opinion, for example owner wants to know what features of a product have been criticised or praised. Following is the example of review posted by a user on a pair of Shoes:

7) I have rated those shoes 4 stars because such a cool pair of shoes, but had a few problems.

8) Order a half size down from your regular fit.

9) Uncomfortable in a few places, but overall not too bad.

10) Arrived 4 days late (DHL's fault), and had a small beige stain next to laces, as well as a black scuff on the white sole.

11) If I was a collector I would have wanted a refund, but I couldn't be bothered to send them back as I had already wait long enough.

12) I would recommend those shoes as they are light weight.

In this situation, the main problem is that what exactly we want to extract from this review. It's easy noticeable that there are variety of opinions provided in this example, sentence (6) express a positive review on shoes while sentence (1) and (3) could be positive or negative. The remaining sentences (4) and (5) are inclined toward negative opinion. Number of opinions in the sentence by user have some targets on which user expressed views. Sentence (4) and (6) are based on the feature of shoes such as quality, shape, fitness whereas sentence (5) is about delivery, nothing related to product. This review example helps to understand the difficulties and challenges in opinion mining or sentiment analysis are directly proportional to deep understanding and require immense data analysis to have precisely analysed opinion. There can be other factors which increases the difficulty in sentiment analysis for e.g. text expressed with irony, sarcasm or negation.

2.2 Sentimental Analysis Classification

Sentimental analysis is classified into multiple sentimental classification techniques. Two of them are popular, first one is machine learning approach and other is lexicon based approach. There is also third classification technique which is known as hybrid approach (show in figure 3) and it uses of the above-mentioned classifications to optimize the solution [13]. Following is the introduction of mentioned approaches.

- Machine Learning Approach: This approach merely depends on text analysis and classification. Text analysis is mainly used for business decision making, for which it require text processing. Initially, it requires some collection of data to train a model, which later serve and help in prediction of new set of data without any sort of labels. Model predicts the unlabelled records by predicting their labelling class. Classes are classified as positive, negative or neutral. Machine learning approach is further divided into following to learning methods.

2. Supervised Learning

2. Unsupervised Learning

Lexicon Approach: In this approach, there is a need to define the dictionary or collection of words and phrases with their synonyms and antonyms. Most common approaches used in lexicon for the collection of words are following two.

3. The Dictionary Based

4. The Corpus Based

- Hybrid Approach: It is most efficient and optimized approach among all, it can identify and detect the emotions from a text. Support Vector Machine Algorithm, which works on a technique to find the best available linear separator between the classes, is required to achieve the goal of hybrid approach.

Sentimental analysis can also be classified in terms of levels and ratings. The classification of these sentiments is based on opinion polarities (positive, negative or neutral) [14]. Figure 3 illustrates the hierarchy and tree structure of classification as discussed earlier.

2.2.1 Classification Levels

This sub chapter explains about the sentimental analysis classification levels. Following are the levels in which sentimental analysis is classified.

5. Sentence Level

It is used as first filter of analysis. In this level of classification, every single sentence is taken under consideration to analyse and express the opinion [15]. Sentence level works on the assumption that there must be single sentiment against one sentence.

This presumption is not necessary for all the sentences in the given collection or document. The most important thing in this classification is to discriminate between biased and non- subjective sentences. Non-subjective sentences provide no information in decision making. Contrarily to this, subjective sentiment provides opinion and detection of those sentences which contains some facts [16]. Sentence level classification provides help to prevent misleading and selecting irrelevant data or sentences. As a result, it is used to increase the efficiency and performance of sentence level sentiment classifier. It is preferred to use when there is need to have more than one opinion in one document. It also provides support to treat sentences differently for special classification. Best scenario of using Sentence Level classification is on conditional and comparative sentences. It is assumed that there is no single strategy available to different sentences or whole text of all the types. In order to improve the accuracy, using combination of different strategies is preferred. It is also preferred to rate opinions in terms of positive or negative opinion, not in terms of good or bad opinions.

6. Document Level

Main objective of this classification is to find out, either whole document has positive or negative opinion. This method considers whole document as a single entity and that is why is not suitable in situation where evaluation of more than one entity requires.

This level is sentiment analysis faces a lot of criticism because of its way of working. As, it is unrealistic due to the fact that there may be many possible opinions in the text. On the other hand, it is useful in situations where some reviews or final statement about the product are required. Other use-case of document level sentiment is scenario news carries some positive or negative opinions and these opinions reflect in terms of buy or sell signals.

7. User Level

This is not a famous or popular level of sentimental analysis but researchers have defined some use-cases for this sort of analysis in a situation where user wants to observe user's network based on the behaviour of the neighbour users [17, 18].

8. Aspect Level

It is also known as feature level or phrase level analysis. It is different from other classification levels in term of method of evaluation. It analyses in such a way that first it finds the target and then discover its opinion. Other the other hand, other classification levels focus on languages units such as sentences, documents and paragraphs. The aim of this classification is not to find the opinion of entities but also their different aspects.

This analysis can be achieved by differentiating polar phrases and defining their sentiments from other [19, 20]. In finance, this level of sentiment is used to find the relationship between detected polar words and other variables, example of which is firm earnings and stock prices [21].

All the models built for product analysis and mining of customer opinions about certain product feature is based on aspect level sentiment [22, 23]. In general, all the words or phrases in aspect level sentiment directed to specified topic or an object.

Figure 4 depicts the process of extracting aspects of reviewed products based on the opinions of customers. Feature selection process is performed to extract feature and based on those feature, trained models categorize the review.

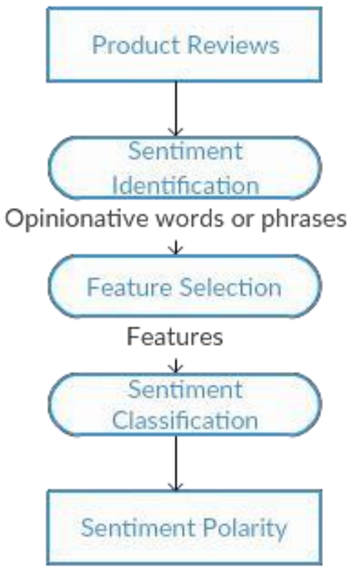


Figure 2 An approach to convert non-grammatical words or phrases

2.2.2 Classification Techniques

In sentiment analysis classification techniques are of core importance and most of work is also done on based on it. The main objective of these techniques is to separate the positive, negative or neutral opinions in document [24]. This sub-chapter mainly focuses on sub-division of machine learning approach and lexicon-based approach, which are discussed previously in chapter 3.2.

Machine learning approach is based on text classification, which is used for forecasting and business decision making by automating the processing text. This approach is divided into supervised and unsupervised learning. In supervised learning approach, initially models are trained using classifiers of document. These trained models or documents have some key features, which have topic related words. Supervised learning is further classified into following.

- Decision Tree: It is used for prediction and are used for classification as well. If record is given with unlabelled or unclassified class label, then compared with decision tree, which is traced from the root to node and find outs the prediction of class against submitted record [25]. Decision tree is popular because it does not require any configuration expertise. Some of the packages used for implementation of decision tree in text classification problems are ID3 and C5.

Linear Classifiers: These techniques are famous due to simplicity. The objective of this technique is to find out number of opinions in provided data and find their polarity by comparing them with list of pre-defined words. Weights are added against words in such a manner, a word with most negative opinion has lowest weight, on the other hand, word with most positive opinion have highest weight. Most popular type of Linear classifiers is Support Vector Machine (SVM) classifiers. [26]

- Rule-Based Classifier: It is same to some extent to technique of decision tree, these techniques are based on rules and feature space [27]. The main distinguish in term of that rule-based classifier allows overlap in the decision tree [26] whereas, decision tree classifier uses hierarchical approach. In this classifier, rules are generated based on different criteria's, such as support and confidence [28].

- Probabilistic Classifier: It is also said to be generative classifiers as it generates a model against each class [29]. It assumes that every class is a part of model. The most widely used probabilistic classifier is Naive Bayer Classifier, which is simplest to implement in any programming language due to the fact that it involves simple mathematics [30]. It works on the principle that each model consists of scattered set of words, frequency of existed words remains same but not the spot. Naive Bayes uses Bayes Theorem, which allows the label to find out the set of features.

Unsupervised learning is used in a situation when it is difficult to create a class- labelled document, which makes it more natural and general as compared to supervised learning. For that purpose, unsupervised learning is implemented on collected unlabelled documents. In case of document clustering analysis, this learning approach is mainly used because it not only relies on already defined class labelled training documents. It is different from supervised learning in such a way it learns by observation and pre-defined models are not submitted to solution. [31]

Another unsupervised approach is Lexicon-based, which uses dictionary. This dictionary consists of list words and phrases, mainly synonyms and antonyms with opinions. Most automated and accepted sentiment word list used for Lexicon-based approach are following.

- Dictionary-Based Approach: It's works on the basic principle that several small set of opinion words are collected manually together to transform it in large collection of text [32]. Every time new word is found, it is added into existing document and this cycle repeats until no unique remains. The major disadvantage of this approach is that it totally depends on large collection of data and it is not possible to enter almost each opinionated word manually created document [33].

- Corpus-Based Approach: Main use of this approach is in scenario where there is need to discover new sentiment word or text from domain of collection in list of already known opinion words and to generate new sentiment lexicon from other [12]. Downside of this approach is that it will work efficiently only in case when collection of all the English or any language words are already present in pre- defined document [34]. It is further divided into statically and semantic approach.

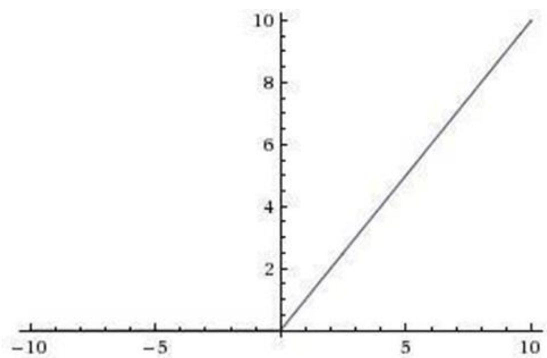
There is also third type of sentimental analysis approach, it is known as Hybrid Approach. The basic working principle of this approach is to find out anticipation from a text with or without affecting associated words. In order to obtain effective identification, Support Vector Machine Algorithm are also used in this approach. Some of industries like HP, are using mixture of Machine Learning and Lexicon-Based approaches together to create hybrid based approach.

4. ReLu Function

It's a very simple activation function. Suppose input is value X and if X is positive the output will be X otherwise 0. ReLu (rectified linear unit) function is:

Function (X) = max (0, X)

In Figure 10 there is a straight line and it looks like it's a linear function but ReLu is non- linear in reality. The range of ReLu is [0, infinity]. Computationally ReLu is less expensive than other activation functions because it has simple mathematical operation.



2.5 Recurrent Neural Networks

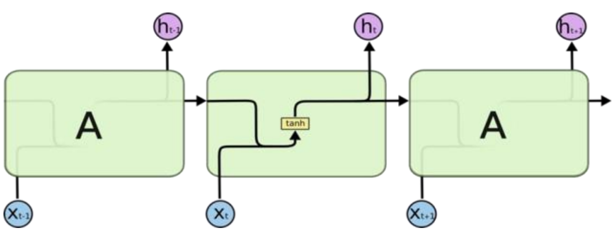
This Chapter explains the difference between human computation and memorizing power and neural networks. In addition to that, it also explains how Recurrent Neural Networks (RNNs) are better than traditional neural networks. It is a human nature that no one start thinking about the situation or problem from the very initial, every now and then. Humans try to sort out the efficient solution of a problem depending on their previous knowledge and understanding. On the other hand, machine neural networks lack this power of taking decision and analysing situations depending on the past information.

In traditional neural networks, it was difficult for machine to memorize the background information. This problem can be elaborated by an example of a movie, consider a machine needs to determine what sort of events could happen next at every scene. It wasnearly impossible for traditional neural networks to deal with such sort of situations. Contrarily to this, Recurrent Neural Networks can handle these issues. In the past few years, RNNs are used widely and almost eliminate use of traditional neural networks. They can be used to solve vast variety of problems, which include, language modelling, image captioning, speech recognition, translation and so on.

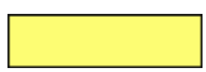
III. THE RNN-LSTM NETWORK PROCESS FOR TEXT ANALYSIS

Long Short-Term Memory network is a special kind of recurrent neural network was proposed by Hochreiter & Schmidhuber [6] as an extension to recurrent neural network. LSTM networks are capable of learning long-term dependencies. It's the default behaviour of LSTM network to remember information for a long period.

A recurrent neural network has a chain of repeating modules of neural network. shows the structure of typical RNN. Each block in the figure is called a module.



Following notations will used to understand:-

- Neural Network Layer 
- Pointwise Operation 
- Vector Transfer 
- Concatenate 
- Copy 

Neural Network

In terms of computer science, a neural network is an artificial nervous system for receiving, processing and transmitting the information. Collection of neurons with synapses8 which connects them is called a neural network. There are three different types of layers in a neural network:

- Input Layer:
- Input fed to network through this layer.
- Hidden Layer:
- This layer processes the input taken from input layer and there can be multiple hidden layers.
- Output Layer:
- This layer produces the processed data.

Figure 6 illustrates the connection between those layers. Circle represents the neurons and the line connecting them represents synapses

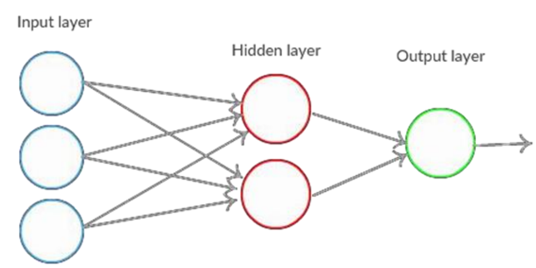


Figure 7 Connection between layers in neural network

Input layer:

Input layer presents a pattern to neural network. This layer only deals with input data. Each neuron in input layer should represent an independent variable that has some effect on the output of network [36].

Hidden layer:

Hidden layer is also a combination of neurons and also has activation function. This layer is also known as middle layer. The main job of hidden layer is to extract important features from data fed by previous layers or layer. There can be multiple hidden layers in network depends upon the complexity of problem. For example, if data can be separated linearly then there is no need to use the hidden layer as activation function can be implemented directly on input layer. If a problem need complex decisions then we can use more than one hidden layer. It's not sure that increasing the number of hidden layers will result in high accuracy.

At some extent accuracy becomes constant or falls if an extra layer has been added. Number of neurons also effect the accuracy result. If number of neurons are less than complexity level of problem then there will be few neurons in the network to detect the signal from complicated data. Similarly, if there are excessive amount of neuron used then over-fitting (explained in section 4.4.2) may occur.

Output Layer:

Output layer collects and produces the results in a way that it has been designed to produce. Typically output layer make predictions for classes.

Neural network training means calibrating the weights and calibration is achieved by repeating forward propagation and backward propagation.

IV. THE RESULT AND ANALYSIS

DATASET AND FRAMEWORK

This chapter contains the information about the framework and dataset, used to train Deep Learning models and the text representations, and also for experimentation in this thesis.

Framework

For this sentiment analysis task, Keras1 has been used for modelling the DL2 (deep learning) models. Keras is a programming framework for deep learning and its written in Python programming language. It is a minimalistic library with a focus on fast experimentation and simplifies the process of building applications based on deep learning. Keras can run on top of Theano or TensorFlow, both of them allows running computations over GPU's. Theano is a library for fast numerical computation. It's a compiler for mathematical computation in Python and was developed by MILA group at University of Montreal, Canada. TensorFlow was created by Google to replace Theano. These two libraries are quite similar but TensorFlow has tools to support Reinforcement learning. Reinforcement learning is a type of machine learning which allow machines to automatically determine the ideal behaviour in a specific context.

Scikit-learn is another library which provides a range of unsupervised and supervised learning algorithms via a consistent interface in Python. But keras library is handier than scikit-Learn. It gives freedom to define our own designed machine learning models, rather than pre-defined ones. We can run Keras on top of TensorFlow. As Google is putting efforts in making TensorFlow the fastest, so this way we can get those benefits. Currently TensorFlow is a scalable (deep learning) engine in the industry.

Combination of Keras and TensorFlow3 has been used for sentiment analysis task. Another package Gensim4 which has been used for word vector handling. Gensim is a Python library which is designed to extract semantic topics from documents. Algorithms used in gensim library are unsupervised and this library is designed to process unstructured, raw text data. Gensim is an exceptionally optimized, yet additionally very specific, library for doing tasks related to text data. It offers a simple, surprisingly efficient AI-approach to handle raw texts and it is based on SNN5 (Shallow Neural Network).

IMDB Dataset

This dataset contains a collection of 50,000 polar movie reviews. Labelled as either negative or positive. Negative reviews hold fewer stars than five stars whereas positive ones were rated with more than six stars. This IMDB data6 has been taken from Stanford University. Researchers in Stanford University collected IMDB data and performed sentiment analysis on that. They achieved 88.89% accuracy. Now the dataset is properly divided, 25,000 for training and 25,000 for testing.

Results:

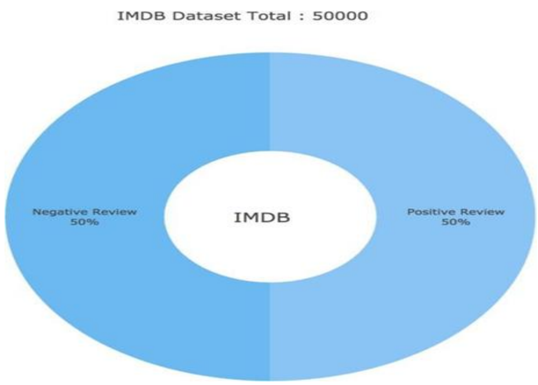
This section presents the results obtained from sentiment analysis. Experiment was conducted on IMDB dataset.

Performance on IMDB Dataset:

As IMDB datasets exhibit very balanced distributions among the positive and negative reviews in Figure 31. Total number of samples are 50,000 from which

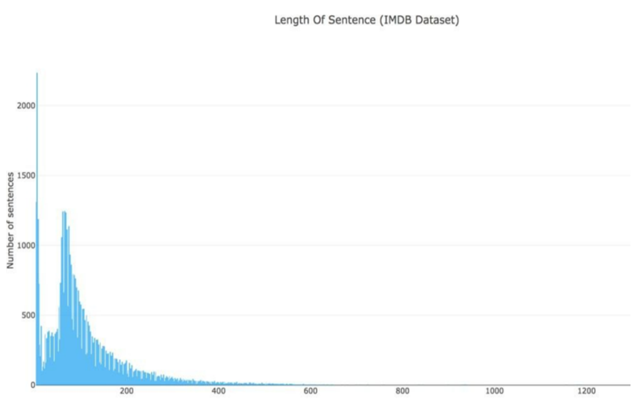
25,000 were positive and negative. This dataset was labelled and modified by Stanford university [9]

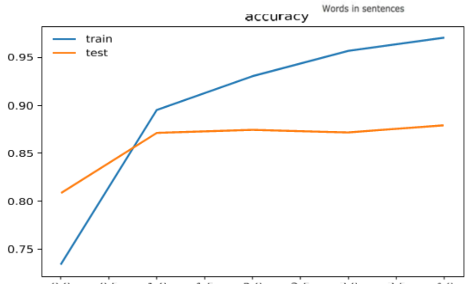
28 Loss is calculated on validation and training and it tells, how well the model is performing.



No. of positive and negative reviews in IMDB dataset

The bars for very short sentences are high which means dataset contains a lot of small sentences. But the frequency of medium sized sentences is very reasonable as in our case a paragraph containing 20 to 200 words is good which can help to analyse the history remembering capability of model. In this data, we have also very huge paragraphs containing max 1400 words but they are less in quantity.





1	1
1	0
1	1
0	0
0	0

Model 96 IMDB 250 1400

REFERENCES

[1] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), 2451-2471.

[2] Graves, Alex. Supervised sequence labelling with re- current neural networks. Vol. 385. Springer, 2012.

[3] A. Graves, A. Mohamed, and G. Hinton. Speech recog- nition with deep recurrent neural networks, ICASSP, 2013. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, and A. Ng, Deep Speech:Scaling up end-to-end speech recognition, in arXiv:1412.5567, 2014.

[4]. Hochreiter, S., & Schmidhuber, J. (1997). Long short- term memory. Neural computation, 9(8), 1735-1780.

[5] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analy- sis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)

[6] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, S. Khudanpur. Recurrent neural network based lan- guage model, In: Proceedings of Interspeech, 2010.

[7] Pascanu, R. and Bengio, Y. (2012). On the difficulty of training recurrent neural networks. Technical Re- port arXiv:1211.5063, Universite de Montreal

[8] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013b. Recursive deep mod- els for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1631–1642, Stroudsburg,

PA, October. Asso- ciation for Computational Linguistics.