International Research Journal of Commerce and Law
Volume 5 Issue 10, October 2018 ISSN: 2349-705X Impact Factor: 4.616
Journal Homepage: http://ijmr.net.in, Email: irjmss@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal

Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data

Prof. Zol Ramdas Madhukar, Prof. Shende SachinSantosh, Prof. Pansare Rajashree Balasaheb, Prof. Dhaygude Tejashri Mohan, Prof.Pandi M.

Dattakala Shikshan Sanstha "Dattakala Group of Institution" Swami-Chincholi, Daund, Pune, Maharashtra 413130. India.

***Abstract*—** In this paper a comparison between different deep learning methods used for sentiment analysis of Twitter data. In this domain, deep learning (DL) techniques, which accommodate at the same time to the solution of a wide range of problems, gained confidence among researchers. There are two types of neural networks are utilized, convolution neural networks (CNN), which are especially performance in the area of image processing and recurrent neural networks (RNN) which are applied with Accomplishment in natural language processing (NLP) tasks. In this study we evaluate and compare combinations of CNN and a category of RNN the long short term memory (LSTM) networks. There are Various tests and combinations are applied and best scoring values for each model are compared in terms of their performance. This research contributes to the region of sentiment analysis by analyzing the performances, advantages and limitations of the above methods with an evaluation procedure under a single testing framework with the same dataset and computing environment. Most of the researches are concentrating on obtaining sentiment features by analyzing lexical and syntactic features. These features are expressed clearly through sentiment words, emoticons, exclamation marks, and so on. Here we introduce a word embedding method obtained by unsupervised learning based on large twitter corpora, this method using latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets. These words are coupled with n-grams features and word sentiment polarity score features to form a sentiment feature set of tweets. The feature set is integrated into a deep convolution neural network for training and predicting sentiment classification labels. We tentatively compare the performance of our model with the baseline model that is a word n-grams model on five Twitter data sets, the results indicate that our model performs civilize on the accuracy and F1-measure for twitter sentiment classification.

Keywords— ***Twitter, word embedding models ,sentiment analysis, deep learning, convolutional Neural networks, LSTM,.***

I. INTRODUCTION

In recent years, there is increase in the use of social media, sentiment analysis gained popularity among a wide range of people with different interests and motivations. As users of all over the world have the chance to express their opinion about different subjects related with education, culture,Commercial products, politics, travel or subjects of general interest, extracting knowledge from those data became a subject of great significance and importance. In addition information about users' visited sites, purchasing preferences etc., knowing their feelings as they are expressed by their messages in various platforms, turned out to be an important element for the estimation of people's opinion about a particular subject. A very common method is to classify the polarity of a text in terms of user's satisfaction, dissatisfaction or neutrality. The polarity can differs in terms of number of levels from positive to negative but in general it denotes the feelings of a text varying from a happy to an unhappy mode. The approaches used for sentiments analysis are numerous and are based on different methods of natural language processing and machine learning techniques for extracting adequate features and classifying text in appropriate polarity labels. Since some years, deep learning methods have obtain, various deep neural networks were utilized on the field with success. Particularly, the neural networks and LSTM networks proved to be tentative for sentiment analysis tasks. Different studies displayed their effectiveness only or in integration between them. In the region of natural language processing, among the methods that are used for extracting features from words, Word2Vec and the global vectors for word representation (GloVe) are the most popular ones.

Text Sentiment analysis is an automatic process to determining whether a text segment contains objective or opinionated content, and it can furthermore determine the text's sentiment polarity. The goal of Twitter sentiment classification is to automatically determine whether a tweet's sentiment polarity is negative or positive.

II. LITERATURE SURVEY

With the amplification and popularity of social media and various platforms allowing people to express their opinion upon different subjects, sentiment analysis and opinion mining became a subject that attracted the attention of researchers worldwide. In a work published in 2008, the authors described the various methods that were used until that day. In the last years deep neural networks proved to be particularly preferment in sentiment analysis tasks. Among them, convolution neural networks and recurrent neural networks were widely implemented because CNN respond very well to the dimensionality reduction problem and a category of RNN the LSTM networks handle with success temporal or sequential data. In the pioneer works presented in and the authors demonstrated that CNN architectures can be utilized with success for sentence classification. Moreover it was demonstrated that CNN perform slightly better than traditional methods. In the efficiency of RNNwas demonstrated as they outperformed the state of the art methods and in an implementation of CNN and LSTM networks was presented, showing the luminous benefit of using jointly these two neural networks.

Twitter is one of the most influencing social media platforms which serve as an information sharing medium in countries all over the world now a days. Therefore, extracting public opinion from tweets about various subjects, measuring the influence of different events or classifying sentiments became a subject of great interest. The sentiment analysis of early works were using different methods for extracting features based mainly on bi-grams, unigrams, POS specific polarity features and were utilizing machine learning classifiers like the Bayesian networks or support vector machines. Most existing studies can be divided into supervised methods and lexicon-based methods to twitter sentiment analysis. Supervised methods are based on training classifiers (such as Naive Bayes, Support Vector Machine, Random Forest) using various combinations of features such as Part-Of-Speech (POS) tags, word N-grams, and tweet context information features, such as hash tags, retweets, emoticon, capital words etc. Lexicon-based methods determine the overall sentiment tendency of a given text by utilizing pre-established lexicons of words weighted with their sentiment orientations, such as SentiWordNet.

These methods rely on the presence of lexical or syntactical features that explicitly express the sentiment information. Though, in a lot of cases, the sentiment of a tweet is implicitly associated with the semantics of its context. In this work, we present semantic feature for sentiment analysis, which is word vector contextual representation of a word in tweet, which can capture the deep and implicit semantic relation information in the words of tweets.

III. METHODOLOGY

. In this section we present the dataset, the word embedding models with their configurations, and the different deep neural network configurations that are used in this study

A. Architecture

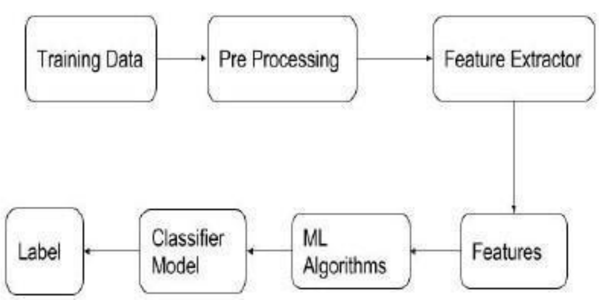

Fig. 1Methodology followed by the system

We use different feature sets and machine learning classifiers to determine the best combination for sentiment analysis of twitter. We also experiment with various pre- processing steps like - punctuations, emoticons, twitter specific terms and stemming. We investigated the following features - unigrams, bigrams, trigrams and negation detection. We finally train our classifier using various Machine-learning algorithms - Naive Bayes, Decision Trees and Maximum Entropy. We present a new feature vector for classifying the tweets as positive, negative and extract Peoples' opinion about products.

B. Dataset and preprocessing

A different datasets was used based on three datasets used in SemEval competitions. More specifically, the SemEval2014 Task9-SubTask B full data, the SemEval2016 full data Task4 and the SemEval2017 development data were used forming a total of around 32.000 tweets. They consist of a body of 662.000 words with a vocabulary of around 10.000 words. In order to increase the system's performance during training the next step was to process the tweets. For this reason, first thing is an extra preprocessing task was carried out to eliminate and innovate some characters. This task enclosed the turnabout of all letters to lowercase, the removal of some special characters and emoticons or the tagging of urls.

C. Word Embedding

The word embedding models used in this study were theWord2Vec, and GloVe. The Word2Vec model was utilized to create 25- dimensional word vectors based on the dataset described before. The configuration of Word2Vec was done by using the CBOW model. Additionally, words that appeared less than five times were discarded. Finally, the maximum skip length between words was set to 5. GloVe was utilized with its pertained word vectors. They are also 25-dimensional vectors and were created from 2 billion tweets, which constitute a significantly larger training dataset than the dataset extracted from SemVal data.

$$v_i' = \frac{v_i - v_{min}}{v_{max} - v_{min}}$$

1) Sentence Vector

The sentence vectors are created after concatenating the word vectors of a tweet in order to form a unique vector. After experiencing with various lengths we created sentences with a length of 40 words. As tweets vary in length, in case that a tweet has more words, the extra words were removed. When they were less than 40 the words of the tweet were repeated until the desired size was achieved. An alternative method is to use zero padding in order to fill the missing words in a sentence. In the approach followed in this work, zero padding was used only in case of words that were not present in the vocabulary.

2) *Sentence Region*

A supplementary approach in word embedding is to divide the word vectors of a sentence in regions, in an effort to preserve information in a sentence and long-distance dependency across sentences during the prediction process. The division is done with the punctuation marks existing on a sentence. In the current configuration each region is composed of 10 words and a sentence has eight regions. In case of missing words or regions, zero padding is applied.

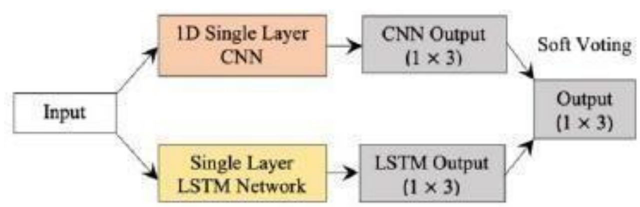

Fig. 2. Regional structure of a sentence

D. Neural Network

The neural network configurations that are proposed for the evaluation of twitter data are based on CNN and LSTM networks. Additionally, in one case a SVM classifier is used. All the networks were tested with both non-regional and regional datasets. In total, eight network configurations are proposed. As mentioned above, RCNN and GRU networks are not utilized because in our experiments they had very similar performance with CNN and LSTM networks correspondingly. All networks were trained with 300 epochs and used sigmoid activation function.

1)Single CNN Network

In this network a single 1-dimensional CNN layer is used. This configuration where the sentence vector is convolved with 12 kernels ith size 1 × 3 (from our tests it performed better when compared with other kernel configurations). The max pooling layer has a size of 1 × 3.

The CNN parameters will be the same for the following CNN configurations. Finally, a 3-dimensional output predicts the polarity in terms of positive, negative or neutral

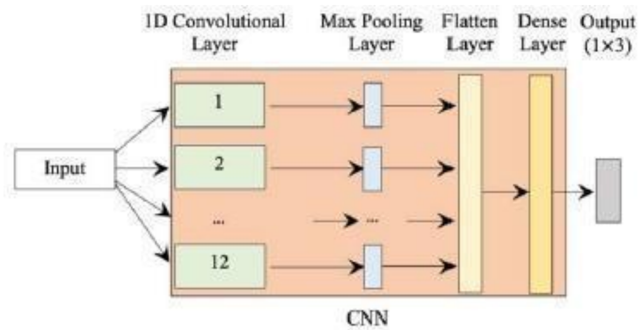

Fig. 3. CNN configuration with one layer and a 3-dimensional output for

2)Single LSTM network

In this configuration a single LSTM layer is used with a dropout of 20%. The output is again 1 × 3 in order to predict the polarity (positive, neutral or negative).

3) Individual CNN and LSTM networks

The aim of this configuration is to take the outputs of individual CNN and LSTM networks and evaluate together their results. A soft voting based on the outputs of the networks decides about the prediction answer. The structure of this configuration where the CNN and the LSTM networks have the same settings as in the two previousconfigurations (for CNN 12 kernels with size 1 × 3 and a max pooling layer with a size of 1 × 3).

The final prediction answer is given after soft voting calculated from the network outputs.

IV RESULT

This section presents the performance results of the previous network configurations in terms of Accuracy, Precision Recall, and F-measure (F1) as described in the following equations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2)$$

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

$$Recall = \frac{TP}{TP+FN} \qquad (4)$$

$$F\_Measure = \frac{2*Recall*Precision}{Recall+PRecision} \qquad (5)$$

In the above equation TP is the true positive, TN is the true negative, FP the false positive, FN the false negative.

Table I shows using CNN and LSTM networks with Wor2Vec and GloVE word embedding systems correspondingly with the performance results of the proposed combinations. First, we can observe that system performance increase of almost all configurations (5%-7%) by utilizing the GloVe. The reason behind it lies to the fact that with Word2Vec the vectorization of words has been made with a relatively small training dataset, around 32.000 tweets compared to the pertained word vectors made with GloVe that used a significantly larger training dataset. The second observation is that using multiple CNN with LSTM networks instead of simple configurations increases the performance of the system, independently of the word embedding system (3%-6%). We can observe that the configurations (2) and (4) have almost always the best performance when compared with the other configurations. A third observation is that separating the text input into regions in most cases doesn’t really improve the performance of a configuration (1%-2%). Concerning the use of SVM classifier instead of a soft- voting procedure it can be seen that it gives a slightly worse performance.

TABLE I. Sentiment prediction of different combination of CNN and LSTM networks with Word2vec word embedding system with no-regional and regional settings from a set of around 32.000 tweets

Network Model	**Embedding word system**				
	Type	***Recall***	***Prec***	***F1***	***Acc***
1.Single CNN network	N-R R	0.33 0.32	0.35 0.34	0.33 0.33	0.49 0.51
2.Single LSTM network	N-R R	0.43 0.44	0.51 0.49	0.39 0.39	0.51 0.50
3.Individual CNN and LSTM networks	N-R R	0.43 0.46	0.47 0.52	0.37 0.42	0.50 0.52
4.Individual CNN and LSTM Networks with SVM classifier	N-R R	0.45 0.42	0.46 0.54	0.43 0.38	0.49 0.51

N-R: Non-Regional, R: Regional word embedding

V CONCLUSION

In this paper various configurations of deep learning methods based on CNN and LSTM networks are tested for sentiment analysis in Twitter data. This appraisal gave fractional but similar results with the state of the art methods, thus allowing expressing credible conclusions about the different setups. The relatively low performance of these systems showed the limitations of CNN and LSTM networks on the field. Concerning their configuration, it was observed that when CNN and LSTM networks are combined together they perform better than when used alone. This is due to the effective dimensionality reduction process of CNN's and the preservation of word dependencies when using LSTM networks. Moreover, using multiple CNN and LSTM networks increases the performance of the system. The difference in accuracy performance between different datasets demonstrates that, as expected, having an appropriate dataset is the key element for increasing the performance of such systems. Consequently, it looks like spending more time and effort in order to create good training sets presents more advantages rather than experimenting with different combinations or settings for CNN and LSTM networks configurations. To summarize, the contribution of this paper is that it allowed to evaluate different deep neural network configurations and experimented with two different word embedding systems under a single dataset and evaluation framework allowing to shed more light on their advantages and limitations.

REFERENCES

[1] W. G. Hatcher, W. Yu: Survey of Deep Learning: Platforms, Applications, and Emerging Research TrendsJ. Clerk Maxwell, A Treatise on Electricity and

MaNg-Rn: Neotins-Rmeg,ion3arl,dR: Reedg.io,navl owlo.rd2em. bOedxdifnogrd Clarendon, 1892, pp.68– 73.

[2] Z. Jianqiang et al.: Deep Convolution Neural Networks for Twitter Sentiment Analysis.

[3] 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

[4] Massa Baali and Nada Ghneim "Emotion analysis of Arabic tweets using deep learning approach".

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741.

[6] International Research Journal of Engineering and Technology (IRJET) "sentiment analysis in twitter" .

[7] J. Spencer and G. Uchyigit, "Sentiment or: Sentiment Analysis of Twitter Data," Second Joint Conference on Lexicon and Computational Semantics. Brighton: University of Brighton, 2013.

[8] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," International Journal, vol. 2, no. 6, 2012.

[9] A. K. Jose, N. Bhatia, and S. Krishna "Twitter Sentiment Analysis". National Institute of Technology Calicut,2012.

[10] Hardt, M., Price, E., Srebro, N., 2016. Equality of opportunity in supervised learning. Advances in neural information processing systems 29