Sentiment Analysis of Tweets by Convolution Neural Network with L1 and L2 Regularization

Abhilasha Rangra(✉), Vivek Kumar Sehgal, and Shailendra Shukla

Department of Computer Science and Engineering,
Jaypee University of Information Technology, Waknaghat, India
abhilasharangra93@gmail.com, vivekseh@ieee.org,
shailendra.shukla@juit.ac.in

Abstract. Twitter data is one of the largest amounts of data where thousands of tweets are generated by the Twitter user. As this text is dynamic and huge so, we can consider it as a big data or a common example of Big data. The biggest challenge in the analysis of this big data is its improvement in the analysis. In this paper, there is an analysis by using semantic features like bigram, tri-gram and allow to learn by convolution neural network L1 and L2 regularization. Regularization is used to overcome the dropout and increase the training accuracy. In our experimental analysis, we demonstrated the effectiveness of a number of tweets in term of accuracy. In the result, we do not obtain any specific pattern but average improvement in the accuracy. For the analysis, we use 10 cross-validations and used to compare the outcome with max-entropy and SVM. Here we also analyze the effect of convolution layer on accuracy and time of execution.

Keywords: CNN · Tweets · Bigdata · Regularization · n-gram

1 Introduction

In the past few years, there was the enormous number of utilization of micro blogging platform like Facebook, Twitter, etc. For the purpose of growth and development, many organizations and companies used to extract the sentiments from the tweet data which also termed as big data. Big data is nothing but a large number of data like Twitter data which we use to analyze the sentiments. We can term tweet data as big data because big data are a set of large data that is used to analyze by using computation methods to demonstrate the trend, association, especially related to human behavior as we are used to exert in Twitter data. In the given paper, we also used to analyze the data from twitter data or we say big data [1, 2]. There are numbers of companies and organization which used to analyze the big data by using several means to identify the opinion of the masses about their product and services [1, 3]. The intention of this paper is to spot some light on the processes, issues, and methods used to analyze the sentiments of big data. This big data contains the number of Twitter messages termed as the tweet, Twitter post, etc. [1, 2]. There is numerous information

A. K. Luhach et al. (Eds.): ICAICR 2018, CCIS 955, pp. 355–365, 2019.
https://doi.org/10.1007/978-981-13-3140-4_32

on the word floating in the big data whether it is related to any organization or any political and social issues. The content in the big data demonstrates real-time events in day to day life or daily routine, these contents are full of social information and temporal attributes. This information, data used to analyze sentiments of human beings, because every person uses to express their view on social sites. After analysis of these data valuable information can be extracted that helps to predict any situation or result easily. Twitter gives fine-grained information about each and every-events, instances, perspectives, etc. [5]. The main intention of sentiment analysis is to evaluate the state of mind of the speaker. In sentiment analysis, most of the work has been done to find sentiment regarding a general topic by taking an assumption that viewer talks about an individual topic. In such reviews, it is easy to analyze the sentiments of the subject [7]. This analysis is used to differentiate the opinion of users or speaker on the basis of its binary polarity. This analysis is done in documents, the single sentence or word for word. The classification of sentiments is done in two ways positive or negative. It is very difficult to make 100% accuracy while analyzing the sentiment, but the main goal of this paper is to provide such approach that assures to make the result as more as possible. We propose CNN for learning purpose which helps to give accuracy in the result. As there are various challenges during analysis of sentiments form big data like to evaluate the data parallel, if we have no previous result Real-time opinion mining becomes the very difficult task, classification of argument statement, the variation of sentiment from person to person or of a person with the passage of time. The sentiment of analysis is done by extracting (REST API), classifying (tokenization, stemming, stop word removal) and last learning (CNN). There is the number of methods, framework, models, etc. given by researchers to the analysis of sentiments of big data in an accurate manner some of the approaches are represented in Sect. 2. The rest of the paper is organized as follows: Sect. 2 gives the detail on the research given by a number of researchers in related work, in Sect. 3 there is a discussion about the proposed work and algorithm, in Sect. 4 there is a discussion about the experiments and result, and at last there is a summary of the conclusion of the whole paper, and the work which is to be done in the future given in Sect. 5.

2 Related Work

To analyze the sentiments of the tweets there is a utilization of lexicon approach and learning based method. Sentiment analysis is performed by collecting big data from the micro blogging site (Twitter) [9]. To execute the analysis, brilliant system with the help of machine learning like Naïve Bayes, Random Frost, or SVM are used [10, 18]. There is a complete description given by [11] about the working during analysis of twitter big data which start from extracting data from social sites and they allow for preprocessing and at last connected with the Alchemy API by Rest Call Method. There is also an evaluation of polarity and accuracy of twitter big data [12] by pre-processing. It is very difficult to analyze the complete paragraph taking adverb, adjective, and verb in a single framework. An AVA framework special for recognizing sentiments in a line, paragraph, etc., where the user has to select topic, line or paragraph, denoted as (t) and any document (d), AVA use to return a score that d used to express (t) [13]. This score expresses in the form of +1 or −1 where +1 means maximally positive and −1 maximally negative. As we discuss above CNN is used in this paper for learning process. Here to extract features from visual and textual modalities by using deep CNN. By using this method in multi-kernel learning classifier we can easily increase the performance to identify the multimodal emotion recognition and sentiment analysis on the different dataset [14]. To analyze the sentiments of the sentences CNN architecture is presented [15] here there is a utilization of 3 pairs of convolution layer and pooling layer in the given architecture. To enhance the given model [15] utilizes parametric rectified linear Unit (PReLU), normalization and Dropout technology.

3 Proposed Work

This section presents the explanation of the proposed methodology and proposed algorithm to analyze the sentiments on big data. Here we give a brief description about the working of CNN and preprocessing methods. Methodology is described in a number of steps with proper flow chart. In algorithm part there is a stepwise discussion about the working of whole methodology and sequence of their execution.

3.1 Preprocessing

In the given paper overall, proposed system is segregated into four steps. In the first step, we use to collect big data and it's pre-processing. In the third step, we use to extract the feature by its classification and at last applied CNN on the labeled feature and classify the outcomes by using soft-max classifier. In Fig. 1, initially, there is a collection of big data with the help of the REST API. Rest API is an application program interface which utilizes an HTTP request to GET, PUT, POST and DELETE data. It is an interface which helps to collect the Twitter data from its site. And in the second step, we use to pre-process the collected tweet for reducing noise. Then apply for tokenization, stop word removal and stemming. In tokenization, we use to make an alphanumeric sequence of collecting data having length character 3 or more. They use to segregate punctuation marks or special characters from the character length and to convert uppercase into lower case [16].

Fig. 1. Pre-processing of big data

In another method [16] given in Fig. 1, there is a removal of the stop word from the sentence and termed as stop word removal. Here it used to remove the word which seems to stop the word. For example, a) "I love Chandigarh but I also love Kinnaur" here, but is utilizing a stop word so this approach used to remove the word 'but'. It also uses to remove extra words or characters, for example "Blassssting day after many days, reeeallyyy enjoooooyinggggg" here there are many extra words who only increasing the data size, this approach used to remove these extra words an left with " Blasting day after many days, really enjoying". It also helps to reduce the data size. This approach uses two methods for removing the words termed as TBRS methods which stand for Random sampling on the basis of a term expressed as:

$$S_z(r) = P_z(r).log_2 \frac{P_{z(r)}}{P_{(r)}} \tag{1}$$

And Mutual interference Method (MI)

$$P(B;M) = \sum_{b \in B} \sum_{m \in M} p(m,b) \log \left[\frac{p(m,b)}{p(m).p(b)} \right] \tag{2}$$

3.2 Semantic Features Extraction

In the third step of Fig. 1, stemming approach [17] is utilized to DE noise the tweets. In this approach, we use to recognize the root word or stem word from the given data. Here we use to remove the plural word like "foods → food and affix word from the data. These are the three methods which are used for the extraction, now we use to classify the given result in the form of n-grams (unigram, bigram and trigram). There are some formulas with help of which we used to classify the features; Fig. 4 represents the flow of feature classification after pre-processing.

In this process, we use to classify the outcome from preprocessed data in term of unigram, bi-gram, and tri-gram. Unigram is utilized to use every individual word as a feature. These features are used for utilization of term frequency, inverse document frequency as the feature value. In unigram n = 1 (Fig. 2).

Fig. 2. Classification of feature

Where,

$$\text{Inverse document frequency} = \log\left[\frac{total\ number\ of\ documents}{number\ of\ documents\ containing - t}\right] \tag{3}$$

Bigrams are used, two consecutive words as features. It is a n-gram where n = 2. It helps to provide the previous token while allowing relation of conditional probability:

$$P\left[S_I|S_0.\ldots\ldots\ldots.S_{I-1}\right] \tag{4}$$

$$P\left(S_1^V\right) = \prod\nolimits_{i=1}^{v} p(S_i|S_{i-1}) \tag{5}$$

and if we talk about the tri-gram having n = 3. It is a group of three consecutive written units like letter, syllables or words. There is one common example of all the three n-grams shown in Table 1.

$$\mathrm{P}(S_i|S_0.\ldots\ldots\ldots.S_{i-1}) \approx \mathrm{P}(S_i|S_{i-2}S_{i-1}) \tag{6}$$

$$\mathrm{P}\left(\mathrm{S}_1^v\right) = \prod\nolimits_{i=1}^{v} p(S_i|S_{i-2}S_{i-1}) \tag{7}$$

Table 1. Accuracy variation in the given three parameters.

	Statistical data for two classes			Statistical data for three classes		
Number of tweets(K)	SVM (Accuracy)	Max-entropy (Accuracy)	CNN (Accuracy)	SVM (Accuracy)	Max-entropy (Accuracy)	CNN (Accuracy)
2	85.62	89.23	92.52	85.23	88.2	90.23
10	86.6	88.42	94.62	87.23	87.98	93.23
40	87.13	86.23	93.23	86.73	85.46	95.67
80	88.23	90.23	96.23	88	88.23	96.23
100	88.47	90.13	97.82	87.98	89.13	97

3.3 Learning Features with Label

After classification of features, we use to apply CNN tool which stands for convolution neural network for learning purpose. Before applying we use to give labeling to the feature by using the correlation method. We use labeled feature for learning like which feature having “positive” sentiments or “negative” sentiments or “neutral” sentiments. CNN is a tool which is very useful or enhanced tool in term of document recognition (LeCun et al. 1998) [18]. It contains any convolutional layers and many fully connected layers. There are two layers named as pooling layer and normalized layer in between the convolution layer. It is a supervised learning algorithm where parameters of various layers are used to learn from back propagation. In the previous year it is very hard to learn large feature set, but after the extension of the computational power of GPU, it is possible to train a deep convolution neural network on the large dataset (Krizhevskhy, Hinton 2012). In our work, there is a use of CNN to learn classified feature in terms of n-grams. There are three main processes in the CNN are pooling, convolution and soft-max model which perform their function on the features. The flow of CNN is represented in Fig. 3.

Fig. 3. Learning process

After labeling the data into a set of vectors which is instantly segregate into a training set and test set. The process of convolution is started from locating word indices into the lower dimensional vector where the distance of the word is directly proportional their sentiments in a layer termed as an embedded layer. In the next layer termed as convolution layer where convolution performs over the embedded words by using various filter sizes which are sliding over 6–7 words at a time. This is a method to understand the activity of words in terms of sentiments in n-grams. After every convolution, there is a utilization to segregate the most significant feature from the set of feature and used to convert them into the feature vector by using the max-pool layer or pooling layer [4]. The combination of convolution and pooling layer is used to make tensor with several shapes and has to create the layer for every individual layer and focus their result of a single big feature vector. Then at last there is a utilization of soft-max model to classify the features as an outcome of pooling layer. The main function of the soft-max model is to optimize the layer by the formula given below [6]. As it is a classifier, it is used to classify the features. [18] L_1 regularization express in Eq. 8.

$$\sum\nolimits_{r=1}^{N} l(\mathrm{f}(\mathrm{Z}_r), \mathrm{I}) + \lambda\left(\sum\nolimits_{r.e=1}^{M+u} L\,\mathrm{f}(\mathrm{Z_r}).\,\mathrm{f}(\mathrm{Z_e}), \mathrm{b}(\mathrm{r}, \mathrm{e})\right) \tag{8}$$

At last to examine the result from learning approach soft max classifier used to classify the outcome in term of accuracy and time execution. L_2 regularization is express in Eq. 9.

Algorithm 1
Step 1: Collection of tweets from Twitter by using REST API.
Step 2: *Pre-process the tweets*
Tokenization → Stop word removal → Stemming
Step 3: *Extraction of processed tweets in the form of n-grams where n = 1 or 2 or 3 respectively. The evaluation is done by using equation 4 and 5 for bigram and 6 and 7 for trigram from the above equation.*
Step 4: *String all the feature.*
Step 5: *Labeled the entire feature by using the correlation method.*
Step6: *These labeled features are applied for learning with the help of CNN.*
$F_r(z) = \sum_{v=1}^{a} B_v^{0,r} \; y^M{}_v(z) + w^{0,r}$, $r= 1,2....K$ *(10)*
Where,
M is the weight of random value
Optimization of feature by using soft-max layer equation (8).
Step 7: *Classification of features and the analysis of accuracy and time of execution of the model using equation (9)*

$$\xi_c : x \rightarrow \sqrt{\frac{2}{\prod} \sum\nolimits_{u \in \Omega_{c-1}} e^{\frac{-1}{W_C^2}\|u-x\|^2} - \xi_{c-1}(u)} \tag{9}$$

4 Experiment

4.1 Experiment Setting

Our experiment is executed on the analysis of sentiments of tweets where we use rest API for the collection of tweets and CNN as a learning approach for the collected data. We used to classify the data in two forms (a) two class which contains positive and negative sentiments (b) three class which contain positive, negative and neutral. We use to analyze the accuracy and execution time for different classes.

4.2 Datasets

To analyze the accuracy in terms of SVM, Max-entropy and CNN we use the dataset from 2 k which having various values for SVM stands for support vector machine, Max-entropy and CNN, similarly for 10 K, 40 K, 80 K and 100 K. These tweets are collected from twitter website with the help of rest API. Tables 1 and 2 shows the statistical data and also provides side-information about the accuracy in SVM, CNN, and max-entropy. Secondly, a number of convolution layers are used to analyze the execution time of CNN where time is considered in seconds. Execution time is represented in the different layer of CNN like 2, 4, 8 and 10 correspond to the accuracy of CNN. Tables 1 and 2 represents the time execution with the accuracy in CNN with the passage of convolution layers.

Table 2. Time taken by the classifier

	Statistical data for two classes		Statistical data for three classes	
Convolution layers	CNN (Accuracy)	Time of execution	CNN (Accuracy)	Time of execution
2	89.92	60	88.23	120
4	92.32	180	92.2	200
8	94.62	240	94.62	320
10	97.82	600	97	800

4.3 Baseline

To represent the analysis we use to compare the accuracy of two classifiers i.e. SVM and Max-entropy and also a deep learning method CNN. There is a comparison of these three approaches to demonstrate the accuracy of the analysis. Here we use to compare the accuracy rate of these three approaches. (Jaynes, 1957) states that max-entropy is the way by which one can explain the data in the best manner with its entropy value. (Berger, 1996) and (pang, 2002) demonstrate the efficiency of max-entropy as compare to NLP and naive Bayes. (Cortes, 1995) and (Renie, 2003) shows the efficiency in classification for selected documents or data. By taking these paper with [19] paper as a base paper, we use to analyze the sentiments of the tweet with statistical data. To generate the statistical data we use to study the behavior of approaches.

4.4 Evaluation Metrics

For recommendation we normally split our measures into three categories Accuracy used to evaluate the overall effectiveness of selected classifier. It is the ratio of Total corrected sentiment (positive + negative) to the complete data used for analysis or classification.

4.5 Results

For the purpose of comparison, we use the number of tweets collected by rest API. The number of tweets varied from 2 K to 100 K for both two class and three class (10-cross-validation). The comparison is done between SVM, CNN, and max-entropy. SVM stands for support vector machine is a state-of-the-art classifier for large data which is used to classify the sentiments of tweets whereas Max-entropy is also used for classification purpose. CNN is described in Sect. 3. Table 1 represents the accuracy variation in the given three parameters.

Table 2 represents the time taken by the classifier (CNN) for attaining the maximum accuracy in the different convolutional layer. There is the various layer of convolution where the value of time varies. Time execution is considered in seconds. Here we use the 100 K number of tweets in 2, 4, 8, 10 layers of convolutions. The given table shows the value of two classes and three classes of (100 K) for CNN execution time in seconds.

4.6 Results Analysis

Tables 1 and 2 demonstrate the values of CNN, SVM, and max-entropy in terms of accuracy and CNN and its execution time with respect to the convolution layers respectively. In Table 1 we use to compare the values of Entropy, SVM, and CNN in terms of accuracy and in Table 2 we use to compare the time with respect of accuracy given by CNN in different layers. The graphical representation of Tables 1 and 2 are shown in Figs. 4 and 5. Figure 4 represents the accuracy of two class and three class given by SVM, max-entropy and CNN for 2 K, 10 K, 40 K, 80 K and 100 K. The graphical result shows the superior performance of CNN in terms of accuracy as compared to the two classifiers in both the cases. Further Fig. 5 represents the execution time of CNN in seconds in term of accuracy. The graphical result demonstrates that with increase in the layer, time execution also increases and accuracy too. We can say that with the increase in this accuracy increase in the layer.

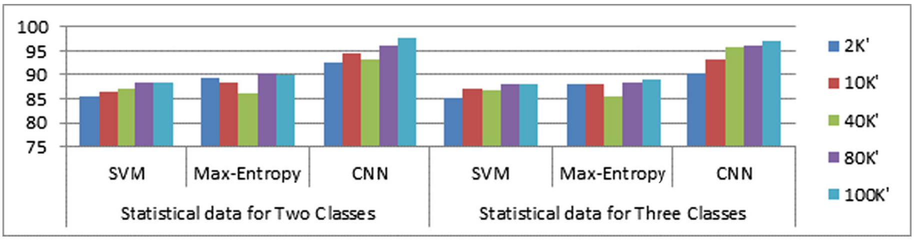

Fig. 4. Comparision of approaches in terms of accuracy

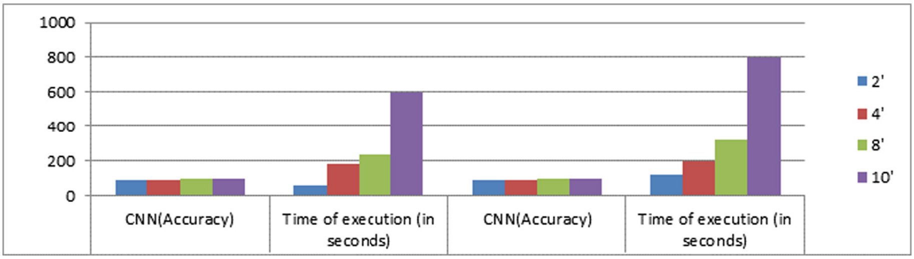

Fig. 5. Comparison of CNN accuary with time (seconds)

5 Conclusion

In the analysis of big data, the important factor is to understand the domain by number of instances. In this paper, experimental analysis is done by using various classifier with different number of instances which starts from 2 K to 100 K. In every case we use 10 cross-validations and generate accuracy. In the experimental result, CNN provides better or effective result as compare to max-entropy and SVM but it also requires more trade–off execution time for two class and three class. The accuracy is near about 97.82 with execution time 600 s and 97 with execution time 800 s for two class and three class respectively. So, if we ignore execution time then we can conclude that CNN shows effective enhancement in accuracy with L1 and L2 regularization from our experimental result in higher number of tweets.

References

1. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224 N Project Report, Stanford **1**, 12 (2009)
2. Agarwal, A., et al.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics (2011)
3. Kouloumpis, E., Wilson, T., Moore, J.D.: Twitter sentiment analysis: the good the bad and the omg! Icwsm **11**(538-541), 164 (2011)
4. Bakliwal, A., et al.: Mining sentiments from tweets. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis (2012)
5. Singh, V., Dubey, S.K.: Opinion mining and analysis: a literature review. In: 2014 IEEE 5th International Conference Confluence The Next Generation Information Technology Summit (Confluence) (2014)
6. Bobicev, V., Sokolova, M., Jafer, Y., Schramm, D.: Learning sentiments from tweets with personal health ınformation. In: Kosseim, L., Inkpen, D. (eds.) AI 2012. LNCS (LNAI), vol. 7310, pp. 37–48. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30353-1_4
7. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: sentiment analysis in Twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (2017)

8. Poursepanj, H., Weissbock, J., Inkpen, D.: uOttawa: system description for SemEval 2013 task 2 sentiment analysis in twitter. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), vol. 2 (2013)
9. Patil, H.P., and Mohammad A.: Applications, issues and challenges in sentiment analysis and opinion mining–a user's perspective. Int. J. Cont. Theory App. **10** (2017)
10. Meral, M., Diri, B.: Sentiment analysis on Twitter. In: 2014 22nd. IEEE Signal Processing and Communications Applications Conference (SIU) (2014)
11. Bhuta, S., et al.: A review of techniques for sentiment analysis Of Twitter data. In: IEEE 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) (2014)
12. Das, T.K., Acharjya, D.P., Patra, M.R.: Opinion mining about a product by analyzing public tweets in Twitter. In: 2014 International Conference on IEEE Computer Communication and Informatics (ICCCI) (2014)
13. Gokulakrishnan, B., et al.: Opinion mining and sentiment analysis on a twitter data stream. In: 2012 International Conference on IEEE Advances in ICT for Emerging Regions (ICTer) (2012)
14. Poria, S., et al.: Convolutional MKL based multimodal emotion recognition and sentiment analysis. In: IEEE 16th International Conference on Data Mining (ICDM) 2016. IEEE (2016)
15. Ouyang, Xi, et al.: Sentiment analysis using convolutional neural network. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM). IEEE (2015)
16. Saif, H., He, Y., Alani, H.: Semantic sentiment analysis of Twitter. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012. LNCS, vol. 7649, pp. 508–524. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35176-1_32
17. Martin, S., Liermann, J., Ney, H.: Algorithms for bigram and trigram word clustering1. Speech Commun. **24**(1), 19–37 (1998)
18. You, Q., et al.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: AAAI (2015)
19. Singhal, P., Bhattacharyya, P.: Sentiment analysis and deep learning: a survey. (2016)