International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

TWITTER OPINION MINING USING NEURAL NETWORKS

DEEPIKA S[1], DEEPIKA S[2], MYVIZHIPRIYA T[2], RAVINDRAN U[3]

1 UG Scholar, Department of CSE, Kingston Engineering College, Vellore-59
2 UG Scholar, Department of CSE, Kingston Engineering College, Vellore-59
3 Asst.Professor, Department of CSE, Kingston Engineering College, Vellore-59

---***---

Abstract - Twitter is one type of social media that is often used. Users use Twitter to convey their tweet to the general public. The number of Twitter users has reached 330 million people worldwide. Besides that, in Twitter there are tweets that can be sentiments. Twitter is one of the most popular online social networking sites where users communicate and interact on various topics. In this paper, Deep CNN Algorithm is used to analysis the sentiments in twitter data. The sentiments are analyzed and labelled as positive and negative labels. These sentiments are visualized as bar graph. This improves the accuracy of the classification.

***Key Words*:** twitter ,mining, negative, positive, netural.

1.INTRODUCTION

To become a Software engineer where I have to use my skills and knowledge for the development of the project, where strong project management and analytical skill will be useful in the delivery of the services to the customer.

Body of Paper

Sentimental analysis is the process of computationally determining the opinion or attitude of the writers as positive, negative or neutral. Data mining is another name for sentimental analysis. In many fields like business, politics and public actions, determining the sentimental analysis is very important. Considering business, it is very useful to understand the customer's feelings in order to develop their company. Next in politics: It can be even be used to predict the election results. There are two ways of classifications and they are (1) machine learning (2) lexicon based approach. In this paper machine learning classifiers are implemented in sentimental analysis and is done in twitter because most of the politicians, famous personalities (even the president of various states) and even general people regularly update their moods in the form of tweets.

Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue.

Data mining is also known as data discovery and knowledge discovery.

The major steps involved in a data mining process are:
Extract, transform and load data into a data warehouse. Store and manage data in a multidimensional databases .Provide data access to business analysts using application software. Present analyzed data in easily understandable forms, such as graphs

Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users.We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

	Machine says yes	Machine says no
Human says yes	tp	fn
Human says no	fp	tn

Table 1: A Typical 2x2 Confusion Matrix

$$\text{Precision}(P) = \frac{tp}{tp+fp} \quad \text{Recall}(R) = \frac{tp}{tp+fn} \quad \text{Accuracy}(A) = \frac{tp+tn}{tp+tn+f+fp+fn}$$

$$F1 = \frac{2.P.R}{P+R} \quad \text{True Rate}(T) = \frac{tp}{tp+fn} \quad \text{False-alarm Rate}(F) = \frac{fp}{tp+fn}$$

2 LITERATURE SURVEY

2.1. OPINION-ASPECT RELATION IN COGNIZING CUSTOMER FEELING VIA REVIEW

Determining a consensus opinion on a product sold online is no longer easy, because assessments have become more and more numerous on the Internet. To address this problem, researchers have used various approaches, such as looking for feelings expressed in the documents and exploring the appearance and syntax of reviews. Aspect-based evaluation is the most important aspect of opinion mining, and researchers are becoming more interested in product aspect extraction; however, more complex algorithms are needed to address this issue precisely with large data sets. This paper introduces a method to extract and summarize product aspects and corresponding opinions from a large number of product reviews in a specific domain. We maximize the accuracy and usefulness of the review summaries by leveraging knowledge about product aspect extraction and providing both an appropriate level of detail and rich representation capabilities. The results show that the proposed system achieves F1-scores of 0.714 for camera reviews and 0.774 for laptop reviews.[1]

2.2. ROLE OF REVIEW NUMERICAL AND TEXTUAL CHARACTERISTIC ON REVIEW HELPFULNESS ACROSS THREE DIFFERENT TYPES OF REVIEW.

Understanding what factors make a helpful online review is critical to increase sales and drive revenue for online retailers. This paper examined the impacts of both reviews' numerical and textual characteristics on review helpfulness across three different review types including comparative, suggestive, and regular reviews. With an analysis of 30338 product reviews collected from Amazon.com, the results indicated that the effects of numerical characteristics of reviews on review helpfulness are stronger for regular reviews than those for suggestive and comparative reviews. The impacts of text sentiment on review helpfulness are more significant for suggestive and

comparative reviews when compared with regular reviews. Moreover, the text complexity of reviews has a significant invert U-shaped relationship with review helpfulness, and the relationships are stronger for regular reviews when compared with suggestive and comparative reviews. Furthermore, text sentiment has a negative effect on review helpfulness, and the effect is stronger for suggestive reviews than that for comparative and regular reviews. Finally, we employ a random forest method to predict review helpfulness based on its numerical and textual characteristics. This paper found that review length is the most helpful factor in predicting the helpfulness of online reviews. Our findings also indicated that the importance of numerical characteristics is greater than that of textual characteristics across three different review types. The theoretical and practical implications of the findings are presented. [2]

2.3. TEXTUAL ANALYSIS FOR ONLINE REVIEW: A POLYMORIZATION TOPIC SENTIMENT MODEL.

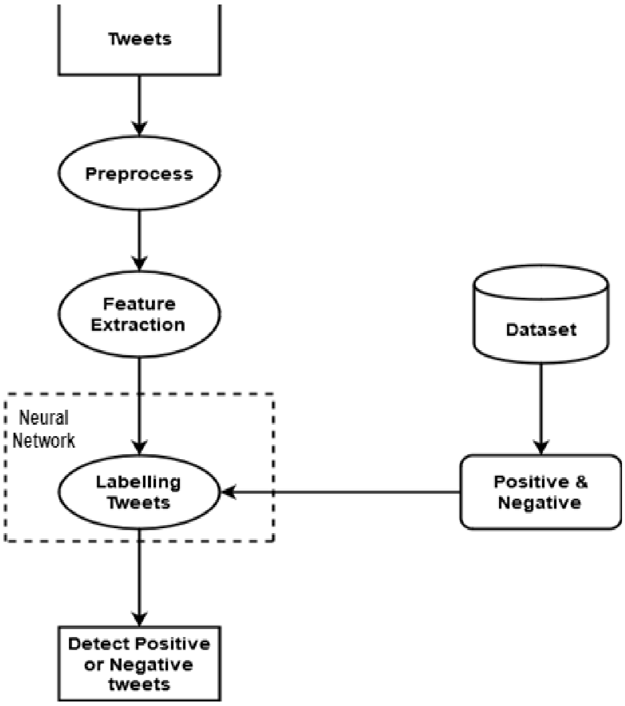
it More and more e-commerce companies realize the importance of analyzing the online reviews of their products. It is believed that online review has a significant impact on the shaping product brand and sales promotion. In this paper, we proposed a polymerization topic sentiment model (PTSM) to conduct textual analysis for online reviews. We applied this model to extract and filter the sentiment information from online reviews. Through integrating this model with machine learning methods, the results showed that the prediction accuracy had improved. Also, the experimental results showed that filtering sentiment topics hidden in the reviews are more important in influencing sales prediction, and the PTSM is more precise than other methods. The findings of this paper contribute to the knowledge that filtering the sentiment topics of online reviews could improve the prediction accuracy. Also, could be applied by e-commerce practitioners as a new technique to conduct analyses of online reviews. [3]

2.4. STATISTICAL AND SENTIMENT ANALYSIS OF CONSUMER PRODUCT REVIEWS.

Big Data commerce has given a big leap to ecommerce .It has opened up the avenues to smarter and informed decision making for large industries as well as the consumers. Online reviews on e-commerce giants like Amazon, Flipkart are one such paradigm which can be used to arrive at more profitable decisions. They are not only beneficial for the consumers but also for the product manufacturers. Online reviews have the potential to provide an insight to the buyers about the product like its quality, performance and recommendations; thereby providing a clear picture of the product to the future buyers. The usefulness of online reviews for manufacturers to realize customer requirements by analyzing helpful reviews is one such unrealized potential. Both positive and negative reviews play a big role in determining the customer requirements and extracting consumer‘s feedback about the product faster. Sentiment Analysis is a computational study to extract subjective information from the text. In this research, data analysis of a large set of online reviews for mobile phones is conducted. We have not only classified the text into positive and negative sentiment but have also included sentiments of anger, anticipation, disgust, fear, joy, sadness, surprise and trust. This delineated classification of reviews is helpful to evaluate the product holistically, enabling better-decision making for consumers. [4]

2.5. HOW CAN OPINION MINING BE USED TO DETECT FAKE PRODUCT REVIEWS?

Positive opinions often mean profits and fame for businesses and individuals. This is a very strong incentive for people to game the system by posting fake opinions or reviews to promote or discredit certain products. Detecting fake reviews is a sub-area of opinion mining. To detect fake reviews, researchers and companies have built detection models using linguistic features (or signals) from the review text content, and meta-data features such as the star rating,user ID of the reviewer, the time when the review was posted, the host IP address, MAC address of the reviewer's computer, the geo-location of the reviewer, etc. Product and sales information is also useful in many cases. For example, if a product is not selling well, but has a large number of positive reviews, these reviews are clearly suspicious.



3.PROPOSED SYSTEM:

Twitter sentiment analysis provides the organizations with the ability to surveying public emotion. In this paper, we introduce the Deep CNN algorithm to analysis the twitter sentiments. The twitter data are fetched using twitter api file. The twitter data are analysed and labelled as positive and negative labelling. Deep CNN algorithm performs both training and testing process to label the sentiments. It results in good accuracy on analyzing the data due to the lack of sentiments analyzing.The amount of training data will be smaller for Deep CNN algorithm

(Working Methodology)

4.ALGORITHM:

Convolutional neural networks or also called convnets are one of the most exciting developments in machine learning in recent years.CNN is a deep ,feed forward artificial neural networks where connections between nodes do not form a cycle and use a variation of multilayer perceptrons designed to require minimal preprocessing . A CNN has hidden layers which are called convolutional layers. When you think of images, a computer has to deal with a two dimensional matrix of numbers and therefore you need some way to detect features in this matrix. These convolutional layers are able to detect edges, corners and other kinds of textures which makes them such a special tool. The convolutional layer consists of multiple filters which are slid across the Text and are able to detect specific features.

LONG TERM SHORT TERM MEMORY:

LSTM or Long Short Term Memory Networks is a specific type of Recurrent Neural Network (RNN) that is very effective in dealing with long sequence data and learning long term dependencies. In this work, we perform sentiment analysis on a GOP Debate Twitter dataset. To speed up training and reduce the computational cost and time, six different parameter reduced slim versions of the LSTM model (slim LSTM) are proposed. We evaluate two of these models on the dataset. The performance of these two LSTM models along with the standard LSTM model is compared. The effect of Bidirectional LSTM Layers is also studied. The work also consists of a study to choose the best architecture, apart from establishing the best set of hyper parameters for different LSTM Models.

5. CONCLUSIONS

In this work, it is found that the classifying the twitter data using the sentiment analysis techniques. This labels the tweets as positive and negative labelling. The classification model is designed in this research work for the hot news detection. The clustering is the labelling of positive and negative tweets using deep CNN algorithm. The performance of the proposed model is analyzed in terms of accuracy, precision and recall.

REFERENCES

[1] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series.," Icwsm, vol. 11, no. 122–129, pp. 1–2, 2010.

[2] M. A. Cabanlit and K. J. Espinosa, "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons," in Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on, 2014, pp. 94–97.

[3] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 1367.

[4] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 625–631.

[5] H. Saif, M. Fernandez, Y. He, and H. Alani, "Evaluation datasets for twitter sentiment analysis," Emot. Sentiment. Soc. Expressive Media, p. 9, 2013.

[6] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," in 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 628–632.

[7] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, vol. 3536, no. c, pp. 1–21, 2017.

[8] R. Sara, R. Alan, N. Preslav, and S. Veselin, "SemEval-2014 Task 9: Sentiment Analysis in Twitter," in Proc. of the 8th International Workshop on Semantic Evaluation, 2014, pp. 73–80.

[9] S. Rosenthal, P. Nakov, S. Kiritchenko, S. Mohammad, A. Ritter, and V. Stoyanov, "Semeval-2015 task 10: Sentiment analysis in twitter," in Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), 2015, pp. 451–463.

[10] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," Proc. 10th Int. Work. Semant. Eval., pp. 1–18, 2016.

[11] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[12] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro, and C. Hervás-Martínez, "Ordinal Regression Methods: Survey and Experimental Study," IEEE Trans. Knowl. Data Eng., vol. 28, no. 1, pp. 127–146, 2016.