

Review Article

CLASSIFYING AND RATIONALIZING TWITTER SENTIMENT THROUGH CLASSIFICATION ALGORITHMS

K.R. SATHISHKUMAR¹, DR.P. VALARMATHIE², DR.K. DINAKARAN³

¹Assistant Professor, Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal, Tamilnadu, India. krsathishkumaar@gmail.com

²Professor, Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil nadu, India. csevalar@gmail.com

³Professor, Department of Information Technology, Saveetha Engineering College, Chennai, Tamil nadu, India. kdinacse@gmail.com

Received: 24.12.2019

Revised: 26.01.2020

Accepted: 28.02.2020

Abstract

Sentiment analysis refers to a classification problem where the main focus is to analyze the sentiment statements which are transferred through the twitter social media. This analysis procedure gives various points of statements from each twit in social media. The huge number of different case of opinions is obtained from the twitter statement and these different opinions are analyzed for predicting the nature of the twit in social media. For a decade, number of different machine learning approaches is used for analyzing and predicting the nature of the twit statements in social media. This paper narrates various conventional twitter sentiment analysis methods with its types.

Keywords: Twitter, Social, Media, Statements, Words, Sentences.

© 2019 by Advance Scientific Research. This is an open-access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.31838/jcr.07.04.24>

INTRODUCTION

The modern media are powerful technique to predict the naturalist of the different kind of persons in world. This modern media is otherwise called as social media where the people in the different regions of the world can send and receive the messages. This advanced social media includes whatsapp, twitter and face book. These kind of social media networks are today more powerful, where the huge number of both educated and uneducated persons are connected through the center hub in the network media. Even though, it is more positive powerful media, sometimes, it is responsible for creating unwanted riots and spreads rumors throughout the world. Hence, the analysis of each transferred statements through this social media is necessary for preventing such kind of unwanted events in world.

Survey on Twitter Sentiment Analysis Framework

Ali Hasan et al. (2018) analyzed the statements of users in social media networks twitter using machine learning languages. The authors extracted policy view factors for each user account from twitter social media and these policy view factors were analyzed by hybrid classification approach which was the integration of Support Vector Machine (SVM) and Naïve Bayes classification algorithms. Figure 1 shows the sentiment analysis from Twitter account. In this paper, the following steps were used for analyzing the twitter account.

Step 1: Data gathering from twitter account.

Step 2: Convert the language into generic language format.

Step 3: Apply preprocessing technique such as removing of URL and symbols.

Step 4: Sentiment analysis using hybrid classification technique.

Step 5: Apply validation technique.

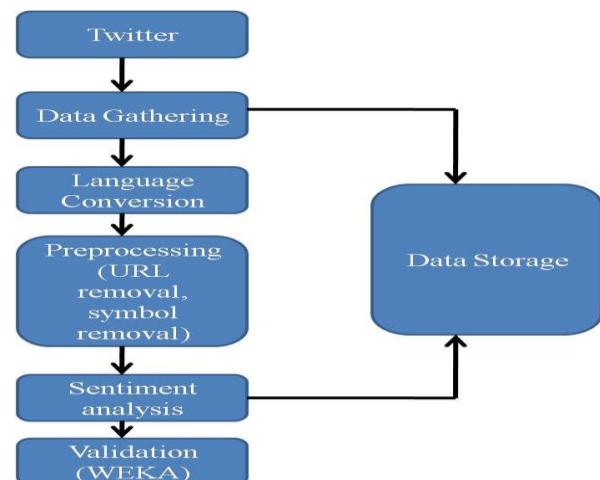


Figure 1: Sentiment analysis from Twitter account

Survey on Sentiment Analysis Using Data Mining Techniques

There are two following conservative sentiment investigation procedure for twitter social media.

- Training and testing scheme;
- Lexicon methodology;

The machine learning approach for sentiment analysis from twitter is classified for improving the level of approximation between various layers. The lexicon learning techniques are also classified into positive and negative approach. Fig.2 details about various conventional strategies for predicting and analyzing the various twits in social media. Further, the machine learning classification approach has been classified into supervised and unsupervised classification techniques.

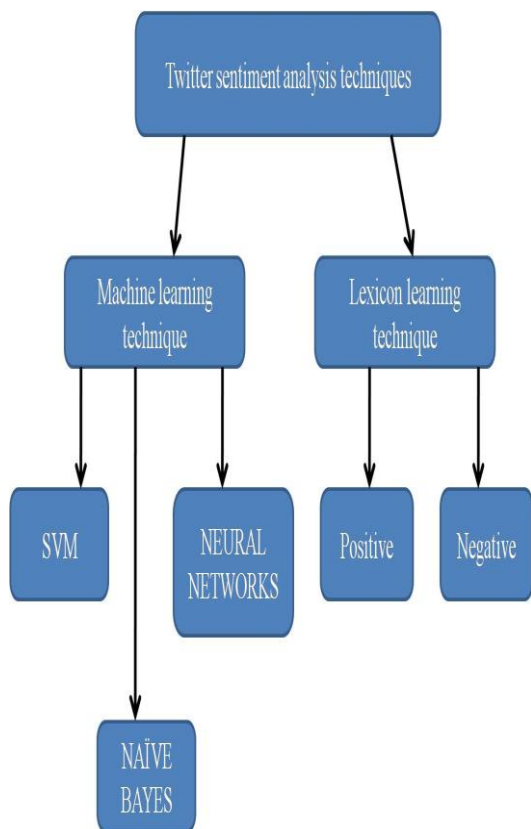


Figure 2: Sentiment Analysis Methods from Twitter

Table 1 shows the twitter analysis methodologies using machine learning technique and lexicon learning technique. The Machine learning technique SVM obtained 86.4% of sentiment analysis report from twitter user account. This classified based technique reaches satisfactory results for obtaining the high level of classification accuracy while it is comparing with various states of conventional approaches.

Table 1: Twitter Sentiment Analysis Methods

Technique	Methodology	Analysis	Accuracy (%)
Machine learning technique	SVM	Twitter	86.4
	Deep learning classifier	Twitter	80.7
Lexicon learning technique	Positive	Twitter	74
	Negative	Twitter	72

Pang et al. (2004) used minimum cut edge technique for analyzing the sentiment from twitter user account. The authors used machine learning classification approach for analyzing the sentiments from twitter account using Support Vector Machine (SVM) classifier. The authors obtained 86.4% of analysis classification report from twitter statement.

Socher, Richard et al. (2013) proposed deep modeling algorithm for analyzing the sentiments from twitter account. The authors used deep neural network cascading and heuristic methods for the case of analyzing the behavior of each twits in social media. The authors achieved 80.7% of accuracy for their classifications from twitter account.

Taboada et al. (2011) used positive and negative approaches for analyzing the sentiment statement from twitter account. The authors obtained 74% of sentiment analysis accuracy for positive approach and also obtained 72% of sentiment analysis accuracy for negative approach. Fig.3 shows the Lexicon method for sentiment analysis.

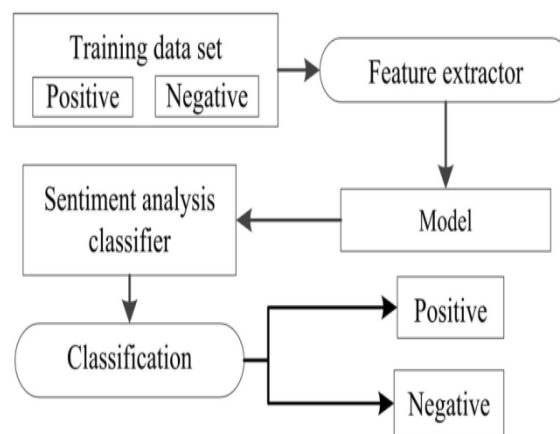


Figure 3: Lexicon Methods for Sentiment Analysis

Survey on Sentiment Analysis Methods

In modern advanced social media, analyzing various properties of each twit method plays an important role in advertisement which improves the economic business in real time world as stated in Bifet and E. Frank (2010). The conventional sentiment analysis method divided into the following categories as stated below.

- Word level analysis.
- Sentence level analysis.
- File level analysis.
- Feature level analysis.

WORD LEVEL ANALYSIS

This method extracts each and every communicated word from twitter account and the extracted words and analyzed for sentiment verification as stated in Agarwal et al. (2011). The word level analysis method can be differentiated as dictionary and corpus techniques, where each method used different planning strategy to predict or analyze the strength of various word analysis techniques. This method produced or attained reasonable results on predicted approaches with respect to various social media works. The corpus methodology used different pattern for computing the main stream of advanced set of twits in each twitter account from the main database in the remote server as stated in Dmitry Davidov et al. (2010). Figure 4 shows the sentiment analysis using corpus dataset.

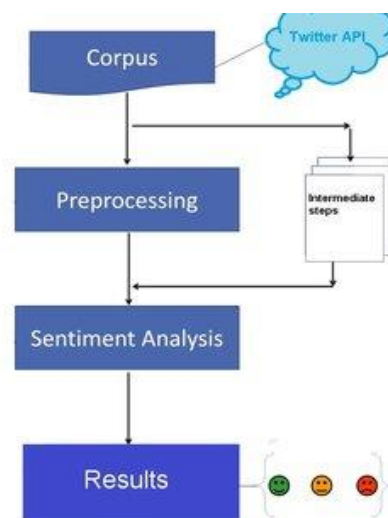


Figure 4: Sentiment analysis using corpus

Sentence Level Analysis

This method extracts each and every communicated sentence from twitter account and the extracted sentences are analyzed for sentiment verification.

File Level Analysis

Paper	Dataset	Technique (precision,%)
Pang et al. [1]	IMDB	NB (81.5), ME (81.0), SVM (82.9)
Turney [2]	Epinions	PMI(66)
Dave et al. [6]	Amazon, CNET	SVM (85.8-87.2), NB (81.9-87.0)
Hu and Liu [4]	Amazon, CNET	Lexicon (84.0)
Abbasi et al. [12]	U.S. & Middle Eastern web forum postings	SVM(95.55)
A. Khan et al. [5]	IMDB, Skytrax, Tripadvisor	Lexicon(86.6)
Zhang et al. [7]	Luce, Yoka	Lexicon(82.62)
Fang et al. [16]	Multi-Domain Sentiment Dataset	ML + Lexicon (66.8)
Zhang et al. [14]	Twitter	ML + Lexicon (85.4)
Mudinas et al. [15]	CNET, IMDB	ML + Lexicon (82.30)

This method extracts file from twitter account and the extracted files are analyzed for sentiment verification.

Feature level analysis

This method extracts features from twitter account and the extracted features are analyzed for sentiment verification. Figure 5 illustrates the different twit analysis streams in social media.

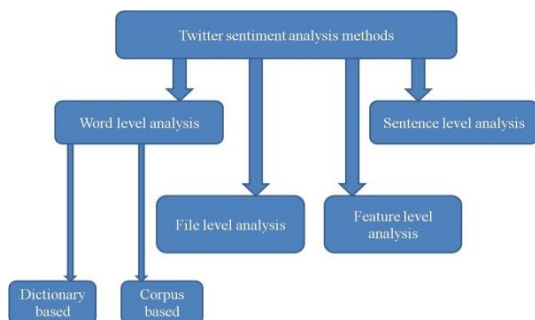


Figure 5: Different Twit Analysis Streams in Social Media

RESULTS AND DISCUSSION

In this paper, the Machine learning technique was utilized to analyze that means classification of twitter sentiment. In that process, the Machine learning technique has given a higher quality or efficiency than the Lexicon learning technique classification. Therefore, the performance of Machine learning technique can be clearly seen through this comparison chart. In Fig.6, 100 records are classified by 85% very clearly. Similarly, other records are categorized more accurately, with 500 records classified as 97% accurate.

CONCLUSIONS

Sentiment Analysis is a technique widely used in text mining. Twitter Sentiment Analysis, therefore means, using advanced text mining techniques in order to analyze various twits in social media networks in the form of multimedia information. The twitter sentiment analysis plays an important role in advertisement which improves the economic business in

real time world. This paper narrates various conventional twitter sentiment analysis methods with its types.

REFERENCES

1. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamsirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts", Mathematical and computational applications, 23,11,2018.
2. Pang, B. and Lee, L. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics. 2004, 271-278.
3. Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment Treebank." Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2013.
4. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M.. "Lexicon based methods for sentiment analysis". Computational linguistics, 2011:37(2), 267-307.
5. Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, 2010, pp. 1-15.
6. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30-38.
7. Dmitry Davidov, Ari Rappoport. "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volume pages 241{249, Beijing, August 2010.
8. Ibrahim, M.; Abdillah, O.; Wicaksono, A.F.; Adriani, M. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14-17 November 2015; pp. 1348-1353.
9. Kaggle. Mean Absolute Error. Available online: <https://www.kaggle.com/wiki/MeanAbsoluteError> (accessed on 26 February 2018).
10. Dokoohaki, N.; Zikou, F.; Gillblad, D.; Matskin, M. Predicting swedish elections with twitter: A case for stochastic link structure analysis. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, France, 25-28 August 2015; pp. 1269-1276.
11. Zemskov, E., Lucas, R., Verin, A.D., Umapathy, N.S.P2Y receptors as regulators of lung endothelial barrier integrity(2011) Journal of Cardiovascular Disease Research, 2 (1), pp. 14-22. DOI: 10.4103/0975-3583.78582