A Comprehensive Study of Sentimental Analysis Methods on Social Media Data

Rajat Kumar Yadu[1], Ragini Shukla[2]

[1]Asst. Prof. Dept. of Computer Science, Mahant Laxminarayan Das College, Raipur (C.G.) India.
[2]Asst. Prof. Dept. of CSIT Dr. C.V. Raman University, Bilaspur (C.G.) India.

ABSTRACT

Sentiments and Opinions are only attributes to express views about attitude, emotions, and sentiments through various social networking sites like twitter, Facebook, Google+ but to categorize accurate positive and negative thoughts of peoples on social media data we have to use various sentimental analysis methods. In this research paper we discuss about various methods related to sentimental analysis and tabulate the accuracy of different techniques and comparing best methods to improve accuracy for social media data. In this research paper we compare various sentimental analysis methods related to classification techniques and create an analysis table for different supervised and unsupervised methods for different social media datasets and accuracy percentage for different techniques. In this research paper we compare various sentimental analysis methods related to classification techniques and create an analysis table for different supervised and unsupervised methods for different social media datasets and accuracy percentage for different techniques. Sentiment analysis relates to the problem of mining the sentiments from online available data and categorizing the opinion expressed by an author towards a particular entity into at most three preset categories: positive, negative and neutral. In this paper, firstly we present the sentiment analysis process to classify highly unstructured data on Twitter. Secondly, we discuss various techniques to carryout sentiment analysis on Twitter data in detail. Moreover, we present the parametric comparison of the discussed techniques based on our identified parameters.

Keywords: Social media, Sentiment Analysis, opinion mining, Decision Tree, Sentiment Methods.

I INTRODUCTION

-Sentiment analysis is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral.

There are various research papers available related to sentimental analysis of social media data where different supervised and unsupervised classification methods used

There are various levels of sentimental analysis such as Document level sentiment analysis, Sentence level sentiment. Sentimental analysis methods for social media data can be categorized can be used in three ways-
From the document, but sometimes nouns or verbs may also express opinion.

II STUDY AREA

In this paper various social media data have used for sentimental analysis where the first step in Supervised Machine learning technique is to collect the training set and then selects the appropriate classifier. Once the classifier is selected, the classifier gets trained using the collected training set.

There various methods used in various sentimental analyses of social media data such as-
(a) **Naive Bayes Classification (NB) -**The statistical Bayesian algorithm characterizes a supervised learning technique, it supposes an originating probabilistic method and it permits us to confine improbability about the method in an honorable way by formative chances of the results. It solves

to correct classify the positive and negative comments by using different tools such as Weka, Matlab and Hadoop and have produced different accuracy percentage.

There are two main approaches for sentiment analysis: machine learning based and lexicon based. Machine learning based approach uses classification technique to classify text. Lexicon based method uses sentiment dictionary with opinion words and match them with the data to determine polarity. They assigns sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary.

predictive and diagnostic and problems and it used to evaluate learning algorithm and provides a useful perspective for understanding.
(b) **Support Vector Machine (SVM) -**In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
(c) **Maximum Entropy algorithm(MEA)-** ME models named as Gibbs, log linear, multinomial logic models , exponential and present a general purpose machine learning technique for prediction and classification which has been productively practical to fields as different as econometrics and computer vision.
(d) **K Nearest Neighbors (KNN) -**In pattern recognition, the k-nearest neighbors algorithm (K-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples.

(e) **10-fold Cross validation**-Cross-validation is a technique to evaluate predictive models by partitioning the original sample into training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples.

(f) **Decision tree** - It is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

III METHODOLOGY

(a) Sentiment Analysis based on Supervised Machine learning technique.
(b) Sentiment Analysis by using Lexicon based Technique.
(c) Sentiment Analysis by combining the above two approaches.

In Supervised Machine learning techniques, two types of data sets are required: training dataset and test data set. An automatic classifier learns the classification factors of the document from the training set and the accuracy in classification can be evaluated using the test set. Various machine learning algorithms are available that can be used very well to classify the documents. The machine learning algorithms like Support Vector Machine (SVM), Naive Bayes (NB) and maximum entropy (ME) are used successfully in many research and they performed well in the sentiment classification.

Table 1
Comparative table for accuracy of various Sentiment Analysis methodsusing Different Techniques

Research Paper	Approach	Dataset	Technique	Accuracy
A Study on Sentiment Analysis on Social Media Using Machine Learning Techniques	Supervised	Twitter dataset	Naive Bayes	53%
			SVM	50%
			KNN	44%
Social media metrics and sentiment analysis to evaluate the effectiveness of social media post	Supervised	Facebook dataset	KNN	82.3%
Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach	Supervised & Unsupervised	Twitter dataset	Naïve Bayes	66.66%
			Hybrid Techniques	76.31%
Classification of Facebook News Feeds and Sentiment Analysis	Supervised	Facebook News Feeds dataset	SVM	97%
			logistic regression	60%
Mining Sentiments from Tweets	Supervised	Google+, Twitter, Myspace dataset	SVM	88%
Sentiment Analysis of Twitter Data: A Survey of Techniques	Supervised	Movie Reviews	SVM	86.40%
		Twitter	Co Training SVM	82.52%
		Stanford Sentiment Treebank	Deep Learning	80.70%

			SVM	82.9%
Analysis of Various Sentiment Classification Techniques	Supervised & Unsupervised	Movie Reviews	MaxEnt	81%
			Naïve Bayes	81.5%
Facebook Posts Text Classification to Improve Information Filtering	Supervised	Facebook News Feeds dataset	SVM	78.9%
			K-NN	55.57%
Sentiment analysis in twitter data using data ytic techniques for lictive modelling	Supervised	Twitter dataset	10- ford cross validation	85%
Sentiment Analysis and Classification of Tweets Using Data Mining	Supervised & Unsupervised	Twitter dataset	Decision Tree	84.66%
			K-NN	50.72%
			Naïve Bayes	64.42%

IV CONCLUSION

In this research paper we compare various sentimental analysis methods related to classification techniques and create an analysis table for different supervised and unsupervised methods for different social media datasets and accuracy percentage for different techniques. The final result of this comprehensive study is that the approx 80% accuracy produced by the support vector machine so SVM is the best suitable sentimental analysis method which can be future used as best classifier for social media data analysis.

REFERENCES

[1] Kumar, P. (2018). A sentiment analysis system for social media using machine learning techniques: Social enablement. Digital Scholarship in the Humanities, IJRAET, Volume-5, Issue -4, 5, pp- 2347 – 2812 doi:10.1093/llc/fqy037.

[2] Poecze, F., Ebster, C., & Strauss, C. (2018). Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. Procedia Computer Science, International Conference on Ambient Systems, Networks and Technologies 130, 660-666. doi:10.1016/j.procs.2018.04.117.

[3] Goyal,S., (2016). Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach, International Journal on Data Science and Technology,Volume2, Issue 5,pp-1-9.

[4] Setty, S., Jadi, R., Shaikh, S., Mattikalli, C., & Mudenagudi, U. (2014). Classification of facebook news feeds and sentiment analysis. International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.2014.6968447.

[5] Bakliwal, A.,(2012). Mining Sentiments from Tweets, International Institute of Information Technology, Volume 7, Issue 12, pp-11-18.

[6] Kharde A.,(2016). Sentiment Analysis of Twitter Data: A Survey of Techniques, International Journal of Computer Applications, Volume 139, Issue 5, pp-5-15.

[7] B., V., & M., B. (2016). Analysis of Various Sentiment Classification Techniques. International Journal of Computer Applications, 140(3), 22-27. doi:10.5120/ijca2016909259.

[8] Benkhelifa, R., & Laallam, F. Z. (2016). Facebook Posts Text Classification to Improve Information Filtering. Proceedings of the 12th International Conference on Web Information Systems and Technologies. doi:10.5220/0005907702020207.

[9] Sulthana, A. R., Jaithunbi, A. K., & Ramesh, L. S. (2018). Sentiment analysis in twitter data using data analytic techniques for predictive modelling. Journal of Physics: Conference Series, 1000, 012130. doi:10.1088/1742-6596/1000/1/012130.

[10] A., V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications, 139(11), 5-15. doi:10.5120/ijca2016908625.

[11] Desai, M., & Mehta, M. A. (2016). Techniques for sentiment analysis of Twitter data: A comprehensive survey. 2016 International Conference on Computing, Communication and Automation (ICCCA). doi:10.1109/ccaa.2016.7813707.

[12] Parveen, H., & Pandey, S. (2016). Sentiment analysis on Twitter Data-set using Naive Bayes algorithm. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). doi:10.1109/icatcct.2016.7912034.

[13] Vairagade, A. S., & Fadnavis, R. A. (2016). Automated content based short text classification for filtering undesired posts on Facebook. 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave). doi:10.1109/startup.2016.7583984

[14] .Kaur, A., & Baghla, S. (2018). Sentiment Analysis of English Tweets Using Data Mining International Journal of Computer Sciences a on Communication, Media, Technology and Design V-6, I-6, pp 25-33.

[15] Shoeb, M., (2017). Sentiment Analysis and Classification of Tweets Using Data Mining, IRJET, Volume: 04 Issue: 12, pp 1471-1474.

[16] Ahmadzadeh, E., & Chan, P. K. (2017), Mining pros and cons of actions from social media for Decision support, IEEE International Conference on Big Data (Big Data).doi:10.1109/bigdata.2017.8258003,V-12,I-14,pp 877-882.

[17] Shoeb, Md. & Ahmed, J. (2017), Sentiment Analysis and Classification of Tweets Using Data Mining,International research journal of engineering and technology (IRJET),V-04,I-12,pp 1471-1474.

[18] Atzmueller, M. (2012), mining social media: Key players, sentiments, and communities. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, V-2, I-5, pp 411-419.doi:10.1002/widm.1069.

[19] Bifet, A.(2013), Mining Big data in Real time ,Informatics ,V-37,I-1, pp 15-20.

[20] Bratawisnu, M. K., Giri, R. R., & Rinaldi, R. (2017), Association perception customer feedback with text network analysis in social media (case study on internet banking BRI, BCA, Mandiri in Indonesia), International Conference on Information and Communication Technology (ICoIC7). doi:10.1109/icoict.2017.8074694,V-10,I-04,pp 1-6.

[21] Barbier, G., & Liu, H. (2011), Data Mining in Social Media. Social Network Data Analytics, pp 327-352. doi:10.1007/978-1-4419-8462-3_12.

[22] Dunren, C. & Mejdl & S. and Zhiyong, P. (2013), From Big Data to Big Data Mining Challenges, Issues, and Opportunities, Published in the Proc. Of International Conference On Database Systems for Advanced Applications Organized & Published by Springer held in Suzhou,V-7827, issue April 22 ,pp 1-15.

[23] Daniel, B. K. (2016), Overview of Big Data and Analytics in Higher Education. Big Data and Learning Analytics in Higher Education published in Springer international journal Switzerland, V-28, I-08, pp1-4, doi: 10.1007/978-3-319-06520-5_1.

[24] Devakunchari, R. Valliyammai, C. (2016), Big Social Data Analytics: Opportunities, Challenges and Implications on Society in International Conference

[25] Felt, M. (2016), Social media and the social sciences: How researchers employ Big Data Analytics. Big Data & Society, V-1, I-15, pp 1-15, doi: 10.1177/2053951716645828.