Research Journal of Computer and Information Technology Sciences _____ISSN 2320 – 6527
Vol. **9(1),** 1-5, June **(2021)** *Res. J. Computer and IT Sci.*

Short Review Paper

A comprehensive survey on various sentiment analysis techniques by machine learning

Somen Roy[1*], Jyothi Pillai[2], Partha Roy[3] and S.P. Shukla[4]
[1]Bhilai Institute of Technology Durg, CG, India
[2]Deptt.of Information Technology, Bhilai Institute of Technology Durg, CG, India
[3]Deptt of Computer Science and Engg, Bhilai Institute of Technology Durg, CG, India
[4]Deptt. of Electrical Engg, Bhilai Institute of Technology Durg, CG, India
somenroy2018@gmail.com

Available online at: www.isca.in
Received th August 2020, revised th November 2020, accepted th December 2020

Abstract

In this survey paper, the general sentiment polarity in reactions to news articles, blogs forum are the main source due to any feeling is known as a sentiment which is also known as opinion mining. The sentiments of individuals toward particular elements are analyzed in this approach. To gather sentiment information, social media is the best-known source. A platform that is accessed socially by various users to post their views is known as blogs. For sentiment analysis, different approaches are used with machine learning and deep learning techniques where supervised and unsupervised algorithms are frequently used for optimization techniques according to their domain by the researchers. In recent years, it has been demonstrated that deep learning models are a promising solution to complex challenges. This paper reviews the latest comparative studies show that have employed traditional machine learning approaches and advance deep learning approaches also able to solve large and complex datasets efficiently in such a way to obtain positive and negative polarity for correctly analysis of output. Different methods used to analysis of sentiment and their approaches provided various outcomes in terms of model performance Finally, a comparative study has been conducted on the experimental results obtained for the die rent models and input features applied by the Authors.

Keywords: Sentiment analysis, classification, regression, machine learning, deep learning.

Introduction

Currently, sentiment analysis is a topic of great interest and development of computer since is a field of research in AI And machine learning and it has many practical applications like health, movie, education, banking share market, sports, any online, offline product, etc. sentiment analysis it is computationally identifying and categorized opinion from a piece of text and image. Companies use sentiment analysis to automatically analyze survey responses, online product review, market prediction through social media comments, they prefer to get customer feedback and their opinion which is effective insights for marketing research about their brands, product, functionality, cost, and services. The customer opinion is an essential part of the prediction of the related item or product, positive and negative opinion is very important for the prediction of sentiment so correctly analyze both features by the various method is the main objective for the success of any project. In this survey paper, a comparative study is conducted for experimental results and authors' approach for sentiment analysis and various techniques used in different fields.

The use of sentiment analysis and machine learning principles in a combined manner to analyze the feelings of users on the Twitter platform the information is extracted from Twitter and therefore the relation between the stock exchange movements of an organization moreover because the sentiment of the tweets are analyzed here. The users are motivated to take a position within the stocks of a corporation when the tweets available on Twitter are positive which could lead to increasing the stock value of the organization in the market. There's a right away relation seen amongst the tweets present by the public and the rise and fall of stock prices[1].

The objective of developing a system that can help in extracting useful information from information available within the application by using classification methods the dataset that has Multilanguage tweets isn't very easy to be handled by the sentimental classification. there's no such technique proposed which may help in handling the multi-language data during this paper, Naïve Bayes and Maximum Entropy classifiers are combined to get one algorithm it seen through the results achieved that the proposed technique has provided better results to other existing approaches[2]. Sentiment classification for mobile reviews they used different classifiers like Naive Bayes, DT (decision Tree), KNN (K Nearest Neighbour), SVM (Support Vector Machine), Random Forest and Fusion approach for in the same dataset to sentiment analysis and their outcomes

shows that SVM based method is overcome to all or any existing approaches[3].

As sentiment classification is finished by various techniques and their approaches is to seek out which combination is best to use within the model building for that they review plenty of paper for sentiment analysis their reviews shows that the among all learning method the SVM based classification gave the higher result[4]. The detection of fraudulent transactions in the master card has been studied using machine learning. First, many standard models are used like NB, SVM, and DL in practical evaluation they have to analyze different outcomes as classifier performance which is almost the same as existing approaches when datasets are used with hybrid models like Ada Boost and majority voting are used. To live performance, it shows that even within the presence of noise, robust performance is obtainable by the majority voting method as compare to other used existing approaches[5].

Within the applications of computing and document classification, various forms of learning methods are utilized. There are various supervised classification methods amongst which any is used within sentiment analysis method because it then a part of text classification problem a classifier that comprises of Bayes theorem is thought as Naïve-Bayes classifier Logistic regression and M. Entropy it's a straightforward probabilistic classifier within it's assumed that the presence or absence of 1 feature during a document isn't reliable on the other feature present within the Document. To produce document classification, Support vector machines (SVM) have been proved to be the foremost efficient technique that will be used for linear classification. Authors are also focused on that dataset that has no label and not required training for that they go with unsupervised techniques for their research[6].

The authors have used a different method for their research and they got outcomes accordingly. it seems that the machine learning approach is the best choice for sentiment analysis and supervised learning is frequently used by the researcher for classification and regression analysis also they have a better result with less manual work and computational time compared to another method. Existing approaches determine the following one of the approaches to develop their model and used learning classifier according to their research requirements[7].

Deep-learning-based techniques (namely CNN, RNN, and LSTM) were reviewed and compared with each other in the context of sentiment analysis probably a better idea to choose for text classification with very large and complex datasets[8].

Sentiment analysis is that the automated process of analyzing text data and classifying opinions into three reasonably sentiment classes positive, negative or neutral. Usually, the machine learning approach is utilized by the researchers and reviews indicate that this research is predicated on machine learning approach has also categories in 3 ways i. Supervise learning, ii. Unsupervised learning, iii. Reinforcement learning.

Sentiment analysis can be done in three ways all existing approaches are based on three categories for the Analysis of sentiment which can understand by broadly categorize as blow.

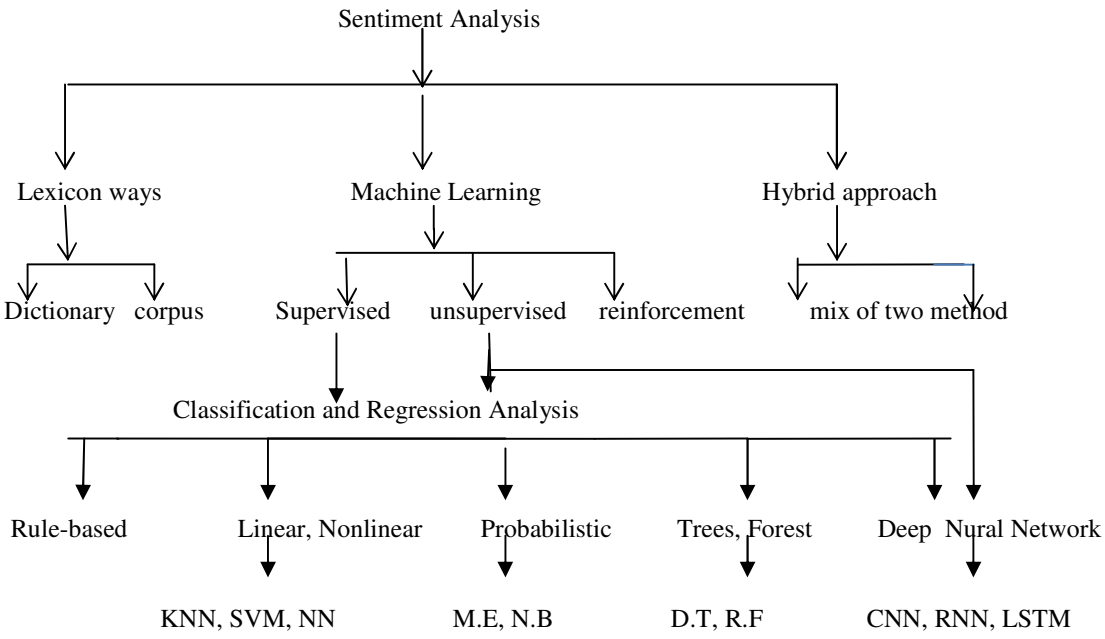

Figure-1: Sentiment Analysis Technique[9,8].

Literature Survey

V.S. Pagolu, et. al proposed[1] the utilization of sentiment analysis and supervised machine learning principles in a combined manner to analyze the sentiments of users on the Twitter platform. The data is extracted from Twitter and the relation between the stock market movements of a company as well as the sentiment of the tweets are analyzed here. The users are motivated to invest in the stocks of a company is a direct relation seen amongst the tweets present by the public with the rise and fall of stock prices.

The objective of developing a system that will help in extracting the useful information from raw data available in the application by using classification methods. This extracted data, which is mainly present in Twitter microblogging applications, has various sentiments present within it that need to be analyzed to analyze the views of users. The dataset which has Multilanguage tweets is not very easy to be handled by the sentimental classification. There is no such technique proposed which can help in handling the multi-language data. In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm among various algorithms it is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches[2].

M.V. Mäntylä, et.al, proposed[3] an approach that utilizes both text mining and qualitative coding to provide the analysis on a large amount of gathered information by the author. With the advent of technology, sentiment analysis is considered the fastest-growing technique in the area of research such as computer science. Hence, the author in this paper analyses various papers and concluded that area of the sentiment analysis is very vast as it covers stock markets, elections, disasters, medicine, software engineering, and cyber bullying, and many more.

C.C and S. Sehgal et al proposed[4], a novel filter-based probabilistic feature selection method that tried to answer the common question that users have when looking for new techniques to select distinctive features to result in improvement of classification accuracy and reduce processing time as well. As such, it is clear that to reach an optimal performance level, and improve the efficiency of classifiers during analysis, it is advisable to include important features in the prediction and extraction of Sentiment information. These important features can be referred to as relevant features. Researched and proposed a model, using n-gram features, stemming, and feature selection to overcome some Persian language challenges (such as the use of informal words) in Sentiment classification. The researchers acknowledged, according to their findings, that feature selection in Sentiment analysis can improve classifier performance.

A. Tripathy, et.al[6] are focused on that dataset that has no label and not required training for that they go with unsupervised techniques for their research To produce document classification, Support vector machines (SVM) have been proved to be the foremost efficient technique that will be used for linear classification.

N. Rajput, et.al, this paper[7] the sentiment analysis is the technique that is applied to analyze sentiment. The sentiment analysis techniques have various phases which are data collection, data cleaning, and classification. In this paper, various sentiment analysis techniques are review and analyzed in terms of certain parameters. In the future, the SVM-based classification method will be further improved for the sentiment analysis.

Nhan Cach Dang et.al In this paper,[8] Author described the core of deep learning models and related techniques that have been applied to sentiment analysis for social network data. We used word embedding and TF-IDF to transform input data before feeding that data into deep learning models. The architectures of DNN, CNN, and RNN were analyzed and combined with word embedding and TF-IDF to perform sentiment analysis experiments, gives us a broad perspective on applying deep learning models for sentiment analysis, as well as combining these models with text preprocessing techniques.

Qian, J ey.al, in this paper,[10] wether analysis related tweets are user comments about daily wether, proposed work gain useful information about how wether influence peoples mood by analyzing them. This is opinion mining in natural language processing field. propose a method modelling text based on deep learning approach, which can automatically extract feature as for word vactor representation, they incorporate linguistic knowledge into words representation in this method.

S.K. Bhatt et.al, in this paper[11] detection of fraudulent transactions in credit cards has been studied using machine learning. First, many standard models are used such as NB, SVM, and DL in practical evaluation. To evaluate, real-world credit card datasets have been used. Along with standard models, some hybrid models such as Ada Boost and majority voting have been used. To measure performance the MCC metric has been adopted so that outcomes are predicted as true for positive and false for negative outcomes. Majority voting gives the best MCC score as 0.823. From a financial institution, real credit card datasets have also been used to evaluate. The same standard and hybrid models have been used. By using Ada Boost and Majority voting a perfect hybrid approach for their future proposed work.

Comparative Study Findings for sentiment Analysis

The main aim of this research is to find how the machine learning approaches are used by existing research and finding to the analysis of sentiment with a certain parameter used by enhancing the performance in the field of sentiment analysis by machine learning approaches and How traditional method can

be applied with different preprocessing techniques and tuning parameters make difference the outcomes in various domain described by the existing research. One last conclusion derived from the study is the observation that the electiveness of the algorithms depends largely on the characteristics of the datasets, hence the convenience of testing deep learning methods with more datasets to cover a greater diversity of characteristics by the Authors.

Conclusion

In this survey paper authors used the various techniques in their approaches to solving the sentiment analysis problems. In their proposed research sentiment reviews are done by machine learning approaches for the classification, regression and clustering are the appropriate choices for the prediction of the model. A comparative study shows the model's advantage and limitations according to the datasets with certain domains also existing methods are focused on supervised and unsupervised techniques with hybrid approaches shows different outcomes. In this review paper comparative study shows traditional machine learning classifier (N.B, DT, KNN, SVM) used with certain parameters, the SVM-based analysis will be further improved by an advanced approach that will make difference for the sentiment analysis. For large and complex datasets in recent approaches, deep learning algorithms CNN, RNN and LSTM may produce better outcomes. Hybrid approaches were identified as the most widely used models for sentiment polarity analysis according to their data domain. In future work exploring hybrid approaches, where multiple models and techniques are combined to enhance the sentiment classification accuracy achieved by the individual models or techniques, as well as to reduce the computational time and cost.

Table-1: Summary of Comparative Study.

Year	Description and method	Experimental result/Target	Ref.
2012	Sentiment analysis on Twitter data by lexicon with machine learning used NB, MNB, SVM, RN.	The result is different outcomes as the dataset size and quality with existing classifier properties.	5
2015	Classification of sentiment review using machine learning for movie reviews used with NB and SVM.	NB has provided better result with Small data set and SVM is great with much larger dataset both have own advantage and limitations also.	6
2016	Sentiment analysis of share market Paper is based on regression analysis of Share market dataset by supervised regression algorithm used.	Based on various experiential results achieved at the end, it is seen that the proposed technique provides better results compare to existing.	1
2017	In this paper sentiment analysis of mobile review by supervised learning. Used classification algorithm and compare with a proposed approach.	According to comparison with the supervised single algorithm, the fusion approach is better with approx. 90 % is the accuracy.	4
2019	Analysis of various techniques of sentiment in terms of certain parameters with machine learning approach.	In the future, the SVM based classification method will further be improved for the sentiment analysis with their dual and kernel properties.	7
2020	Sentiment Analysis based on deep learning: A Comparative Study	DNN approach overcomes the existing traditional approaches.	8
2018	Sentiment Analysis Model on Weather-Related Tweets using (NB, SVM, DNN) supervised approach.	A neural network with incremental machine learning which is deep neural network gave better accuracy than existing.	10
2020	Credit card fraud detection by Ada-Boost and Majority of boating as a hybrid approach.	It's a combination of two methods in one approach which overcome the other existing approach Majority of Voting gives the best MCC score.	11
2019	Target Sentiment Analysis Model with Naïve Bayes and Support Vector Machine for Product Review Classification.	Best result outcome according to NB with Small dataset is good but it is Probabilistic however SVM provides the better result with a large dataset and there is more option to improve.	12
2019	Sentiment Analysis of Keenly Intellective Smart Phone Product Review Utilizing SVM Classification Technique.	Among all Traditional algorithm SVM and their hyper parameter Kernel, Gama, and C are the used for complex problem with better result.	13
2020	Sentiment Analysis and Clinical Analytics. In Innovation in Health by NLP.	NLP Deep learning with embeddings gave better result in accuracy and convergence.	14

References

1. V. Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi(2016), "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", IEEE.

2. Ankit Kumar Soni (2017), "Multi-Lingual Sentiment Analysis of Twitter data by using classification Algorithms" in IEEE.

3. Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila(2018), "The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers", Volume 27, Pages 16-32,

4. Chhaya Chauhan and Smriti Sehgal (2017), "Sentiment analysis on product Reviews" International Conference on Computing Communication and Automation'vol. 10 pp 1-8, IEEE.

5. B. Balakrishnan et al(2012)., Opinion Mining and Sentiment Analysis on a Twitter Data Stream, The IEEE International Conference on Advances in ICT for Emerging Regions', Pp 182-188.

6. A.Tripathy, Ankit Agrawal and Santanu Kumar Rath(2015)."Classification of Sentimental Reviews Using Machine Learning Techniques", 3rd International Conference on Recent Trends inComputingRTC-2015) Classification of Sentimental Reviews using machine Learning.

7. Neha Rajput and Mrs. Shivani Chauhan(2019), et al Analysis of various techniques of sentiment an International Journal of Computer Science and Mobile Computing, Vol.8 Issue.2.

8. Nhan Cach Dang 1, María N. Moreno-García2, and Fernando De la Prieta 3,(2020) Sentiment Analysis Based Deep Learning: A Comparative Study Electronics 2020, 9, 483; doi:10.3390/electronics9030483.

9. P. Routray, C. K. Swain and S. P. Mishra(2013), "A Survey on Sentiment Analysis," International Journal of Computer Applications 'Volume 76 – No.10, p.0975 – 8887**.**

10. Qian, J.; Niu, Z.; Shi, C.(2018) Sentiment Analysis Model on Weather-Related Tweets with Deep Neural Network. In Proceedings of the 10th International Conference on Machine Learning and Computing, Macau, China 26–28 February 2018; pp. 31–35

11. Sanath Kumar Bhat, K, Sreenath N, Divya BS(2020) Credit Card Fraud Detection using Machine Learning Methods International Journal for Research in Applied Science & Engineering Technology Volume 8 issue

12. Richard rimfire, A.ogutu(2019) Target Sentiment Analysis Model with Naïve Bayes and Support Vector Machine for Product Review Classification International Journal of Computer Science and Information Security (IJCSIS) Vol. 17, No. 7,

13. Dubey. T& Jain. A(2019) 'Sentiment Analysis of Keenly Intellective Smart Phone Product Review Utilizing SVM Classification Technique' 10th ICCCNT 2019 July 6-8, IIT - Kanpur, India IEEE45670.

14. Rajput, A. (2020),Natural Language Processing, Sentiment Analysis, and Clinical Analytics. In Innovation in Health Informatics; Elsevier: Amsterdam, The Netherlands, 2020; pp. 79–9

15. Satapathy, R.; Cambria, E.; Hussain, A(2017). Sentiment Analysis in the Bio-Medical Domain; Springer: Berlin, Germany,

16. Bhavitha, B.; Rodrigues, A.P.; Chiplunkar, N.N(2017). Comparative study of machine learning techniques in sentimental analysis. In Proceedings of the International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 10–11 March 2017; pp. 216–221.

17. T.U. Haque, N. N. Saber, and F. M. Shah(2018), "Sentiment analysis on large scale amazon product reviews," IEEE International Conference on Innovative Research and Development (ICARD).IEEE,

18. Jangid, H.; Singhal, S.; Shah, R.R.; Zimmermann, R(2018) Aspect-Based Financial Sentiment Analysis using Deep Learning. In Proceedings of the Companion of The Web Conference on The Web Conference, Lyon, France, 23–27 April 2018; pp. 1961–1966.