© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 04 | July - September 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

A Survey on Sentiment Analysis of Twitter Data

Sapna Manjhi , M.A. Rizvi

[1] M.tech student, [2]professor
[1]computer technology and science
National Institute of Technical Teachers Training and Research, Bhopal

ABSTRACT
In modern-day web, the online users are terribly active in registering their opinions a couple of product or a service. There are numerous choices like social media, public forums and E-commerce sites for registering the opinions. With the rise of such sources on the online, people andorganizations are progressively victimization such public opinions for creating important selections. Sentiment analysis is that the method of trailing public reviews to extract the hidden positive, negative or neutral sentiments within the statement and thereby creating effective selections. Sentiment analysis involves grouping the offered data, extracting the options, choosing the required options and eventually creating the sentiment classification to attain the opinions. The extracted opinion offers suggestions for purchasers to grasp additional concerning merchandiser service before shopping for. Sentiment analysis is useful for the makers and marketers to gauge their success on new unleash of their merchandise or service in market. Additionally to any or all the higher than benefits, sentiment analysis helps the policy manufacturers to grasp the general public read and to form public friendly policy, through that new government policies may be simply analyzed. This paper present review on different technique and algorithm are applied in sentiment analysis.

KEYWORDS: Twitter, Sentiment Analysis, Microblogging.

I. INTRODUCTION

The growth of social media are increase within the variety of users has made-up the approach for posting one’s opinion in own language. Sentiment analysis in social media is a lively field of analysis within the recent years. As part of the investigation of this language processing system, a model is looked for that can identify the many dimensions of attitudes utilised by social media users at different times. Using the extensive evaluation of sentimental analysis models, the research will be able to identify the user's fashion and strategy on feelings related to a couple of different topics, services, or events, for instance. Models for sentiment analysis are known to pick up on the attitudes and views of users.

An opinion mining technique known as sentiment analysis utilize the natural language processing, content assessment, and computational linguistics to reduce or extract subjective facts from source sources. Ads and customer service are only two examples of how sentiment assessment may be used in a wide range of industries.

II. SENTIMENT ANALYSIS

Text, audio, tweets, and database sources are all examples of sources that may be used for sentiment analysis. Natural Language Processing (NLP) can be used to extract sentiments from these sources. The process of Sentiment analysis categorise views expressed in a piece of writing into "positive, negative, and neutral" categories. Variety of terms are used to describe it, including subjective analysis, opinion mining, and evaluation extraction.

Phrases opinion, feeling, view, and belief are utilised interchangeably, yet there are subtle variances.

- *Opinion*: A conclusion that may be disputed (because different experts have different opinions)
- *View*: viewpoint that is purely subjective.
- *Belief*: acceptance that is deliberately and intellectually assented to.
- *Sentiment*: a person's self-described point of view

III. TWITTER SENTIMENT ANALYSIS

When doing sentiment analysis on tweets, the primary objective is to achieve an accurate classification of the tweets into a number of distinct sentiment classes. This area of study has given rise to a number of different techniques, all of which provide ways to provide training to a model, and then examine it to determine whether or not it is effective. As previously indicated, doing sentiment analysis on Twitter data is difficult. In this section, we explain why this is so:

- **Limited tweet size**: Compact statements are constructed with just 140 characters in hand, resulting in a minimal collection of features.
- **Use of slang**: If an approach is outmoded due of the emergence of slangs, these terms are distinct from English words.
- **Twitter features**: It enables for the usage of hashtags, user references, and URLs. These must be handled in a different way than 9ordinary words.
- **User variety**: The users express their viewpoints in a wide range of methods; some of them switch to a new language in the middle of their statements, while others utilise the same words or symbols many times to communicate their feelings.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 04 | July - September 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

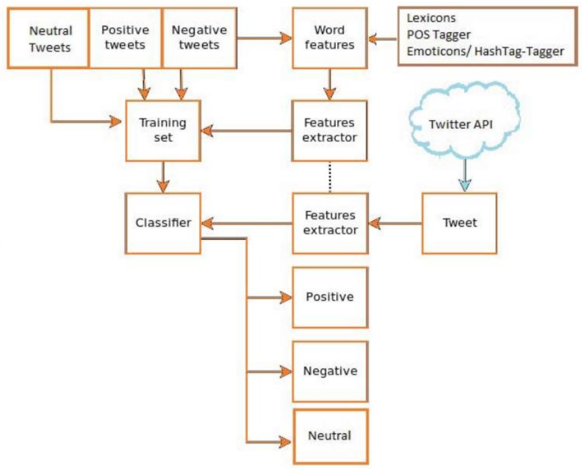

Fig. 1.1 Sentiment Analysis Process

They must be addressed under the area of preprocessing. In addition, we have difficulties in feature extraction when we have fewer features to work with and when we reduce the dimensions of features.

IV. RELATED WORK

"Sentiment Analysis" or Opinion Mining has been subject of a great deal of study. These investigations are focused on identifying the accurate meaning of an electronic message. "Machine learning and semantic orientation" are two of most common categories for these methods. We'll go through the existing research on each of these techniques in this section.

A. *Machine learning*

In order to use machine learning, 2 sets of documents are needed. A set of training and a set of test are available. To begin, machine learning algorithm must be skilled on both supervised and unsupervised learning works (such as classification and forecast, for example) (clustering etc.). During the training phase, the algorithm is taught how to function using a set of predetermined inputs so that it may subsequently be evaluated using undefined inputs. In order for our algorithm to be able to categorise unknown inputs, we must first train it. Sentiment Analysis uses a variety of machine learning techniques. This section goes through a few of them.

It's one of the most successful and basic methods out there. Nave Bayes It's a widely-used text categorization method (Melville [1],Rui [2],Ziqiong [3],Songho [4],Qiang [5] and Smeureanu[6]). The final probability is determined using this method by first calculating prior probability of an entity belonging to class, and then multiplying that prior probability by the likelihood of the entity belonging to the class. The approach is nave because it relies on the erroneous assumption that each word in the text stands alone. This is a more convenient assumption, but the accuracy suffers as a result.

Support Vector Machines (SVMs) are another option (SVM). Text classification using a discriminative classifier is another use for it (Rui [2],Ziqiong [3],Songho [4], and Rudy [7]). Structural risk reduction is the guiding idea of this method. There are two distinct types of training data points, which are determined by the decision criteria or surface. In the training set, the choice is based on support vectors chosen. "Support Vector Machine" (SVM) has number of distinct variations, one of which is a multiclass SVM, which is used for sentiment analysis [8].

The K-Closest Neighbor (KNN) technique [4] identifies text document's K nearest neighbours among the training documents. The process of categorization is carried out based on the similarity score that each class has in relation to its neighbouring documents.

B. *Semantic Orientation*

Unsupervised learning is at the heart of the semantic orientation method. Sentiment data classification doesn't need any training. It is used to determine the word's polarity, which may be either positive or negative. Sentiment analysis was carried out by Kamps[9] using lexical relations. An approach based on semi-supervised learning was presented by Andrea [10], and it included the introduction of a seed set and the subsequent expansion of that set using Word Net. Because words with identical orientation are presumptively of same polarity, this was the assumption. In Chunxu[11], sentiment analysis may be performed on material whose contextual information is unknown. In this procedure, additional relevant materials were utilised to extract necessary contextual information, and then that information was utilsed to determine orientation of the opinion. Unsupervised learning method on the basis of part of speech (pos) pattern presented by Ting-Chun [12]. Sentiment phrases were utilised as queries in search engine, and sentiments were predicted using the result of search. For sentiment analysis, Gang [13] used "TF-IDF (term frequency – inverse document frequency)" weighted. To further stabilise the grouping, they employed K-means clustering on the raw data and a voting method. The papers were divided into two categories based on the results of several process implementations. An easy-to-use lexicon-based method was utilised by Prabhu [14] to recognise and extract emotions from Twitter data.

Table I - Comparison and Summary of different sentiment analysis techniques

S. No.	Technique	Learning Methodology	Advantages	Disadvantages
1	SVM	Supervised	A high degree of accuracy with a low degree of overfitting.	Inability to classify multi-classes computationally costly and slow
2	Naïve bayes	Supervised	It is easier to learn and categorise Insensitive to traits that are not relevant.	Assumes that each feature has its own distinct identity. SVM is more accurate.

			Streaming data is easily handled.	
3	Centroid classifier	Supervised	Low cost of computation data set with a lot of dimensions It's possible to merge numerous characteristics into one	In a class, there is a term dependence I'm very concerned with the training data Numerous characteristics in the feature vector
4	KNN	Supervised	Train with breakneck speed An easy-to-understand system Training data that can withstand the effects of noise. Well-suited for handling enormous amounts of data	Influenced by the K value Complexity of computing is high. Improper qualities might easily mislead him.
5	Winnow classifier	Supervised	Adopting a method based on errors Focused on the interplay between characteristics Weights are only changed while a feature is in use.	SVM is less accurate. On a variety of training datasets, tuning is not consistent.
6	K – means clustering	Unsupervised	Much more rapid than traditional forms of guided instruction Clusters are formed easily and quickly.	The accuracy of this method is less than supervised learning. K's value is hard to anticipate. Doesn't function well with varying densities and densities of clusters.

V. CONCLUSION

There is an overview and comparison of current methodologies in this work. Computer-based algorithms like SVM and naive Bayes are the most accurate and may be considered the baseline for learning. Different applications are already making use of sentiment analysis, and there will be many more in the future. That much is clear from the above explanation. We may draw the conclusion that the more accurate the findings are, the cleaner the data is. It has also been shown that multiple strategies may be coupled to overcome the limitations of each other and produce a superior categorization overall. The classification methods need to be improved further. A number of issues, including the management of implicit product features and the handling of negation, among other things, have not yet been entirely resolved.

VI. REFERENCE

1. Melville, WojciechGryc, "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification", KDD"09, June 28–July 1, 2009, Paris, France.Copyright 2009 ACM 978-1-60558-495-9/09/06.
2. Rui Xia, ChengqingZong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", Information Sciences 181 (2011) 1138–1152.
3. Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment classification of Internet restaurant reviews written in Cantonese", Expert Systems with Applications xxx (2011) xxx–xxx.
4. Songbo Tan, Jin Zhang, "An empirical study of sentiment analysis for chinese documents", Expert Systems with Applications 34 (2008) 2622–2629.
5. Qiang Ye, Ziqiong Zhang, Rob Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", Expert Systems with Applications 36 (2009) 6527–6535.
6. Ion SMEUREANU, Cristian BUCUR, "Applying Supervised Opinion Mining Techniques on OnlineUser Reviews", InformaticaEconomică vol. 16, no. 2/2012.
7. Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach." Journal of Informetrics 3 (2009) 143–157.
8. KaiquanXu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50 (2011) 743–754.
9. Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using wordnet to measure semantic orientation of adjectives", Proceedings of 4th International Conference on Language Resources and Evaluation, pp. 1115-1118, Lisbon, Portugal, 2004.
10. Andrea Esuli and FabrizioSebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of 14th ACM International Conference on Information and Knowledge Management,pp. 617-624, Bremen, Germany, 2005.
11. Chunxu Wu, LingfengShen, "A New Method of Using Contextual Information to Infer the Semantic Orientations of Context Dependent Opinions", 2009 International Conference on Artificial Intelligence and Computational Intelligence.
12. Ting-Chun Peng and Chia-Chun Shih , "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and intelligent Agent Technology JOURNAL OF COMPUTING, VOLUME 2, ISSUE 8, AUGUST 2010, ISSN 2151-9617.
13. Gang Li, Fei Liu, "A Clustering-based Approach on Sentiment Analysis", 2010, 978-1-4244-6793-8/10 ©2010 IEEE.
14. PrabuPalanisamy, VineetYadav, HarshaElchuri, "Serendio: Simple and Practical lexicon based approach to Sentiment Analysis", Serendio Software Pvt Ltd, 2013.

15. Liu, S., Li, F., Li, F., Cheng, X., &Shen, H.. Adaptive cotraining SVM for sentiment classification on tweets. In Proceedings of the 22nd ACMinternational conference on Conference on information &knowledgemanagement (pp. 2079-2088). ACM,2013.
16. Pan S J, Ni X, Sun J T, et al. “Cross-domain sentiment classification viaspectral feature alignment”. Proceedings of the 19th internationalconference on World wide web. ACM, 2010: 751-760.
17. Wan, X..“A Comparative Study of Cross-Lingual SentimentClassification”. In Proceedings of the The 2012 IEEE/WIC/ACMInternational Joint Conferences on Web Intelligence and IntelligentAgent Technology-Volume 01 (pp. 24-31).IEEE Computer Society.2012
18. Socher, Richard, et al. "Recursive deep models for semanticcompositionality over a sentiment Treebank." Proceedings of theConference on Empirical Methods in Natural Language Processing(EMNLP). 2013.
19. Meng, Xinfan, et al. "Cross-lingual mixture model for sentimentclassification." Proceedings of the 50th Annual Meeting of theAssociation for Computational Linguistics Volume 1,2012
20. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., &Stede, M..“Lexiconbasedmethods for sentiment analysis”. Computational linguistics, 2011:37(2), 267-307.