Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.8 No.S2, 2019, pp. 66-69
© The Research Publication, www.trp.org.in

Classification Techniques on Twitter Data: A Review

S. Shafina Banu[1], K. Syed Kousar Niasi[2] and E. Kannan[3]
[1]Research Scholar, Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science & Information Technology,
Jamal Mohamed College, Tiruchirappalli, Tamil Nadu, India
[3]Registrar, VELTECH-Vel Tech Dr. Rangarajan and Dr. Sagunthala R&D Technical University, Chennai, Tamil Nadu, India
E-Mail: sshafinabanu1992@gmail.com

***Abstract* - Data mining is the practice of examining unknown patterns of data according to diverse viewpoints for classification into valuable information, which is composed and gathered in collective areas, such as data warehouses.For effective analysis, data mining algorithms enabling business decision making and other information necessities to eventually cut costs and raise revenue. Sentiment analysis is the method of defining the emotional tone behind a sequence of words, used to gain an accepting of the attitudes, opinions and emotions conveyed within an online mention. Sentiment analysis is tremendously useful in social media observing as it allows us to gain a synopsis of the broader public opinion behind definite topics. The applications of sentiment analysis are extensive and influential. The ability to abstract insights from social data is a practice that is being broadly adopted by organizations across the world. In this paper, we focused on sentiment analysis on the twitter data.**

***Keywords:* Data Mining, Sentiment Analysis, Social Media, Twitter, Emotions**

I. INTRODUCTION

In the previous few years, there has been an enormous development in the use of microblogging platforms such as Twitter[1]. Encouraged by that development, companies and media organizations are gradually looking for ways to mine Twitter for information about what people think and feel about their products and services. Companies such as Twitratr (www.twitrratr.com), tweetfeel (www.tweetfeel.co m) and Social Mention (www.socialmention.com) are just a rare who promote Twitter sentiment analysis as one of their services.Social media such as Weblogs, microblogs, and discussion forums are used day-to-day to precise personal thoughts, which allows researchers to gain valued insight into the opinions of a very huge number of persons, i.e., on a scale that was merely not potential a few years ago. As an outcome, currently, sentiment analysis is usually used to study the public opinion towards persons, events and objects[2].

The identification of sentiment in text is a significant field of study, with social media platforms such as Twitter gathering the attentiveness of researchers in language processing as well as in political science and social science. The task commonly includes identifying whether a piece of text expresses a POSITIVE, a NEGATIVE, or a NEUTRAL sentiment; the sentiment can be general or about a particular topic, e.g., a product, a person, or an event [3].

The birth of Web 2.0 invented a breaking down on the block between the consumers and producers of information. In other words, the web improved from a static container of information into an energetic element in which any user in a very modest manner could circulate any type of information. The information produced by users varies broadly, from publications in blogs or other forums to simple comments on their state of mind in social networks. This enablement of publication has led to the growth of numerous different websites specifying in the publication of users' opinions.

Discovery sources of information and observing their progress on the web is a very difficult task due to the huge number of different sources and the huge volume of texts; each with their individual opinions becoming even more complex when the opinions are not conveyed clearly. These huge amounts of information make it very complex for a human reader to identify and select opinions on the Internet [4]. For this reason, it is essential to improve systems for instinctive retrieval, search, classification and the presentation of points of view. This new discipline, named opinion mining (OM) or sentiment analysis (SA), has risen in order to resolve this difficult problem.

II. RELATED WORK

Barbosa L and Feng J said that using n-grams on tweet data may delay the classification performance because of the huge number of rare words in Twitter[5]. Instead, they suggested using micro blogging features such as re-tweets, hash tags, answers, punctuations, and emoticons. They create that using these features to train the SVMs improves the sentiment classification accuracy by 2.2% related to SVMs trained from unigrams only.

Agarwal *et al.,* [6] discovered the POS features, the lexicon features and the micro blogging features. Apart from basically relating various features, they also designed a tree demonstration of tweets to combine numerous categories of features in one brief representation. Moschitti proposed a partial tree kernel was used to compute the comparison between two trees [7]. They found that the best significant features are those that combine prior polarity of words with their POS tags. All other features only play a minimal role. Additionally, they also indicated that joining unigrams with the greatest set of features out performs the tree kernel-based model and provides about 4% absolute gain over a unigram baseline. Speriosu *et al.,* made a graph that has several micro blogging features such as hash tags and emoticons together with users, tweets, word unigrams and bigrams as its nodes which are associated based on the link presence amongst them (e.g., users are related to tweets they produced; tweets are related to word unigrams that they comprise etc.)[8].They then applied a label propagation technique where sentiment labels were spread from a lesser set of nodes seeded with some primary label information through the graph.

Yang *et al.,* [9] use web-blogs to build corpora for sentiment analysis and use emotion icons given to blog posts as displays of users' mood. The authors applied SVM and CRF learners to categorize sentiments at the sentence level and then examined numerous strategies to determine the complete sentiment of the document. As the outcome, the persuasive strategy is defined by considering the sentiment of the latest sentence of the document as the sentiment at the document level.

III. SOCIAL MEDIA

Social media denotes to the means of interactions between people in which they generate, share, and/or interchange information and ideas in virtual communities and networks. Social media is about conversations, community, relating with the audience and constructing relationships. It is not just a transmission channel or a sales and marketing tool. Authenticity, honesty and open dialogue are important. Social media not only allows you to hear what people say about you, but allows you to reply. Listen first, speak second. Be convincing, useful, related and engaging. Don't be frightened to try new things, but think over your efforts formerly kicking them off[10] is the main objectives of Social Media.
Popular Social Media Tools and Platforms:

1. *Blogs:* A platform for unplanned dialogue and debates on a definite topic or opinion.
2. *Facebook:* The world's biggest social network, with more than 1.55 billion monthly active users (as of the third quarter of 2015). Users make a personal profile; add new users as friends, and interchange messages, comprising status updates. Brands make pages and Facebook users can "like" brands' pages.
3. *Twitter:* A social networking/micro-blogging platform that permits groups and individuals to stay linked through the interchange of short status messages (140 character limit).
4. *YouTube & Vimeo:* Video presenting and viewing websites.
5. *Flickr:* An image and video presenting website and online community. Photos can be united on Facebook and Twitter and additional social networking sites.
6. *Instagram:* A free photo and video sharing app that permits users to apply digital filters, frames and special effects to their photos and then share them on a diversity of social networking sites.
7. *Snap chat:* A mobile app that allows users send videos and photos to friends or to their "story." Snaps disappear after watching or after 24 hours. Presently, we are not permitting separate departments to have Snap chat accounts, but asking that they donate to the Tufts University account.
8. *LinkedIn Groups:* A place where groups of professionals with related areas of interest can share information and partake in conversations.

IV. SENTIMENT ANALYSIS

Sentiment analysis denotes to the class of computational and natural language processing based techniques used to detect, abstract or describe subjective information, such as opinions conveyed in a given piece of text. The key purpose of sentiment analysis is to categorize a writer's attitude towards numerous topics into positive, negative or neutral groups.Sentiment analysis has several applications in different domains comprising; but not limited to business intelligence, politics, sociology, etc. Recent years, it is observed that the initiation of social networking websites, microblogs, wikis and Web applications and so;a unique evolution in user-generated data is poised for sentiment mining. Data such as web-postings, Tweets, videos, etc expresses opinions on numerous topics and events which offer huge chances to study and examine human opinions and sentiment.

Sentiment analysis (sentiment classification, opinion mining, subjectivity analysis, polarity classification, affect analysis, etc.) is the multidisciplinary field of study that compacts with examining people's sentiments, attitudes, emotions and opinions about diverse entities such as individuals, products, organizations, services, companies, events and topics that contains multiple fields such as natural language processing (NLP), information retrieval, computational linguistics, artificial intelligence and machine learning. It is a combination of computational and NLP centered techniques which could be leveraged in order to abstract particular information in a known text unlike factual information, opinions and sentiments are subjective [11].

V. DATA MINING TECHNIQUES

Data mining is extremely effective, so long as it draws upon one or more of these techniques[12]:

1. *Tracking patterns*: One of the supreme simple techniques in data mining is learning to identify patterns in your data sets. This is commonly acknowledgment of some aberration in your data happening at fixed intervals, or an ebb and flow of a definite variable over time.
2. *Classification*: Classification is a more tough data mining technique that powers you to gather numerous attributes collected into apparent groups, which you can

then use to draw further conclusions, or assist some function.

3. *Association*: Association is associated with tracking patterns, but is more particular to dependently related variables. In this case, you'll look for particular events or attributes that are extremely interrelated with another event or attribute;
4. *Outlier Detection*: In many cases, basically identifying the overarching pattern can't give you a perfect understanding of your data set. You also need to be able to recognize anomalies, or outliers in your data.
5. *Clustering:* Clustering is much related to classification, but includes grouping chunks of data composed based on their comparisons.
6. *Regression*: Regression, used mainly as a form of planning and modeling, is used to recognize the probability of a certain variable, given the occurrence of other variables.
7. *Prediction*: Prediction is one of the greatest valuable data mining techniques, later it's used to project the types of data you'll see in the future. In various cases, just identifying and accepting historical trends is sufficient to chart a somewhat exact prediction of what will happen in the future.

VI. COMPARATIVE STUDY OF VARIOUS PROPOSALS

Mashael Saeed Alqhtani *et al*., [13] aims to classify social network to anomaly groups such as: Terrorist and dissident; by analyzing tweets data on the Twitter; then classify an anonymous user's affiliation to these groups. To address this difficult, they first abstract a set of features to describe each group using dissimilar data mining techniques and store these features in the database. Text mining, sentiment analysis, and opinion mining techniques will be used to achieve this extraction. The objective of data extraction is to measure the comparison of selected user tweets with respect to mined features. It will enable to define great percentage of resemblance between the user tweets and group characteristics to uncover his/her affiliation to this group.

MdShoeb*et al*., [14] made an attempt to categorize sentiment analysis for tweets with the help of text mining and data mining techniques. They use three dissimilar classifiers – Decision Tree, K-NN, and Naïve Bayes. All the three classifiers predicted the labels for a dataset. The outcome displays that the accuracy of Decision Tree, K-NN and Naïve Bayes is 84.66%, 50.72%, and 64.42% respectively. The outcome also shows that the precision of Decision Tree, KNN and Naïve Bayes is 95.96%, 90.00%, and 67.08% respectively. They said that Decision Tree classifier is the greatest classifier to be used with social media dataset as it gives the more exact and detailed prediction.

Kirti Huda *et al*., [15] in this research work, pattern based technique is applied for the feature abstraction in which patterns are created from the existing patterns which rise the accuracy of data classification. The proposed algorithm has been applied in python and it has been analyzed that execution time is abridged and accuracy is improved at steady rate. The proposed development is applied in python and it is analyzed that execution time is reduced to 10 percent and accuracy is rise to 20 percent.

Amandeep Kaur *et al*., [16] exploited the fast memory computation framework 'Apache Spark' to abstract live tweets and execute sentiment analysis. The major aim is to offer a method for analyzing sentiment score in noisy twitter streams. This research work reports on the design of a sentiment analysis, mining huge number of tweets. The result categorizes user's perception through tweets into positive and negative. Secondly, they discussed numerous techniques to carryout sentiment analysis on twitter data in detail.

Efstratios Kontopoulos *et al*., [17] suggest the distribution of original ontology-based techniques towards an extra effective sentiment analysis of twitter posts. The novelty of the projected approach is that posts are not basically characterized by a sentiment score, as is the case with machine learning-based classifiers, but instead collect a sentiment grade for each separate notion in the post. The proposed architecture outcomes in a more complete analysis of post opinions about a specific topic were accurate.

TABLE I COMPARATIVE STUDY ON VARIOUS PROPOSED WORKS

Authors	Model	Classifier	Outcome	Accuracy	Dataset
Mashael Saeed Alqhtani *et al*., [13]	Supervised Machine Learning Technique	Naïve Bayes, Decision Trees using C4.8, Random Trees classifiers	Classifies anomaly groups	NB – 87.14% DT – 87.14% RTC – 84.29%	Tweets
MdShoeb *et al*., [14]	Text Mining Techniques	Decision Tree,K-NN, Naïve Bayes	Classifies sentiment analysis for tweets	DT - 95.96%, K-NN - 90.00% NB - 67.08%	Tweets
Kirti Huda *et al*., [15]	Pattern Based Technique	N-gram algorithm & SVM Classifier	Analyze the user behavior	Execution time is reduced to 10 % and Accuracy is rise to 20%	Twitter Data
Amandeep Kaur *et al*.,[16]	'Apache Spark' Framework	Custom Algorithm	User's perception on tweets (positive or negative)	Score: Positive Words: 0 Negative words: -2	Noisy Twitter Streams
Efstratios Kontopoulos *et al*.,[17]	Ontology-based techniques	Machine learning-based classifiers	Analysis of post opinions	71% to 86%	Twitter Posts

VII. CONCLUSION

In this paper, we presented a survey for sentiment analysis on Twitter data. Sentiment analysis is a developing field of data mining used to examine the sentiment of the data, whether it is positive, negative or neutral. Emoticons are extensively used to precise people emotions or opinions. At the present time, social media are playing a chief role in sharing people's knowledge, thoughts, views or ideas on specific topics like political issues, economical problems, social issues, also self-statuses. In social media, people use emoticons to precise their emotions; in such case social media like Twitter, Facebook produce vast amounts of data sets. This paper discussed the role of emoticons on Twitter and the models to define the range of sentiment on the twitter data which was updated by the user on certain range of sentiment, which influences people in wide.

REFERENCES

[1] Efthymios Kouloumpis, Therasa Wilson and Johanna Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*., 2011.

[2] SaraRosenthal, Preslav Nacov, Svetlana kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov,"SemEval-2015Task10: Sentiment Analysis in Twitter", *Proceedings of the 9th International Workshop on Semantic Evaluation., Jun. 2015*.

[3] Sara Rosenthal, Noura Farra, and Preslav Nakov, "SemEval-2017Task4: Sentiment Analysis in Twitter", *Proceedings of the 11th International Workshop on Semantic Evaluations,*Aug. 2017.

[4] Eugenie Martinez Camara, M. Teresa Martin Valdivia, L. Alfonso Urena Lopez, and Arturo Montejo Raez, "Sentiment analysis in Twitter", Oct. 2012.

[5] L. Barbosa, and J.Feng, "Robust sentiment detection on twitter from biased and noisy data", *Proceedings of COLING.*, pp. 36-44, 2010.

[6] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data",*Proc. ACL 2011 Workshop on Languages in Social Media.*, pp. 30–38, 2011.

[7] A. Moschitti, "Efficient Convolution Kernels for Dependency and Constituent Syntactic Trees",*M. (eds.) ECML 2006. LNCS (LNAI)*, Vol. 4212, pp. 318–329,2006.

[8] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter polarity classification with label propagation over lexical links and the follower graph", *Proceedings of the EMNLP First Workshop on Unsupervised Learning in NLP*., pp. 53–63, 2011.

[9] ChanghuaYang, KevinHsin-YihLin and Hsin-His Chen, "Emotion classification using web blog corpora", *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*., pp. 275–278, 2007.

[10] [Online] Available at: https://communications.tufts.edu/marketing-and-branding/social-media-overview/

[11] GhazalehBeigi, Xia Hu, Ross Maciejewski and Huan Liu, "An Overview of Sentiment Analysis in Social Media and its Applications in Disaster Relief", *Springer, 2016*.

[12] Larry Alton. [Online] Available at: https://www.datasciencecentral .com/profiles/blogs/the-7-most-important-data-mining-techniques, 2017

[13] Mashael Saeed Alqhtani and M. Rizwan Jameel Qureshi, "Data mining approach for classifying twitter's users", *International Journal of Computer Engineering & Technology (IJCET)*, Vol. 8, No.5, Oct. 2017.

[14] Md. Shoeb and Jawed Ahmed, "Sentiment Analysis and Classification of Tweets Using Data Mining", *International Research Journal of Engineering and Technology (IRJET)*, Dec. 2017.

[15] Kirti Huda, Md. Tabrez Nafis and Neshat Karim Shaukat "Classification Technique for Sentiment Analysis of Twitter Data", *International Journal of Advanced Research in Computer Science*, Vol. 8, No.5, Jun. 2017.

[16] Amandeep Kaur, Deepesh Khaneja, Khushboo Vyas, and Ranjit Singh Saini, "Sentiment Analysis on Twitter using Apache Spark", *Advanced Topics in Computer Systems.*, Oct. 2017.

[17] Efstratios Kontopoulos, Christos Berberidis, The ologos Dergiades, and Nick Bassiliades, "Ontology-based sentiment analysis on twitter posts", *Expert Systems with Applications,*Vol. 40, No.10, pp. 4065-4074, Aug. 2013.