Twitter Sentiment Analysis Using Textual Information and Diffusion Patterns

Prof. R Y Totare[1], Aishwarya AherGawli[2], Abhijeet Girase[3], Ishwari Tale[4], Ayushi Khanbarad[5]

[1,2,3,4,5]All India Shri Shivaji Memorial Society, Institute of Information Technology, Pune, 411011

Abstract
Billions of individuals daily post their ideas on social media platforms like twitter. As a brief and straightforward means to express human feelings, the tweet. Therefore, we concentrated on sentiment analysis of Twitter data in this article. The majority of the sentiment analysis tools for Twitter now in use mostly focus on the textual content of tweets and make an effort to function well when dealing with brief and confusing tweets. According to recent studies, the polarities of Twitter tweets and the ways in which emotions are expressed on the platform are closely related. This research focuses on how to combine the textual content of Twitter tweets and sentiment propagation models in order to increase the effectiveness of sentiment analysis in Twitter data. The suggested approach explores the dispersion of feelings by examining a phenomena known as the inversion of feelings and identifies numerous intriguing aspects of the reversal of feelings in order to do this. In order to anticipate the polarities of the feelings conveyed in Twitter messages, we propose using random forest machine learning. We do this by considering the connections between the textual content of Twitter messages and the sentiment-dissemination patterns. This research is the first of its kind to apply sentiment dissemination models to enhance Twitter's sentiment analysis, as far as we are aware. Numerous tests on a real-world dataset demonstrate that, in contrast to cutting-edge text-based analyticsmethods.

Keywords- Text Mining, Machine learning, Sentiment analysis, sentiment diffusion, Twitter.

I. Introduction

Twitter has altered and changed how people get information from the people or organisations that interest them., a well-known microblogging service around the world. Tweets are status update messages that users can post on Twitter to let their followers know what they are thinking, doing, or what is going on in the world. Users can also communicate with one another by reacting to or reposting their tweets. Twitter has grown to be one of the world's largest online social media platforms since it was founded in 2008. Due to the expanding amount of data coming from Twitter and its many applications, the polarity of the emotions expressed by users mined in Twitter messages has become a popular research topic.For instance, several systems have been developed to propose tactics for political elections by analysing the political views of Twitter users on political parties and candidates. Twitter sentiment analysis is another tool used by businesses to quickly and efficiently track consumer attitude towards their goods and services.

In order to do this study, a number of words or tweets were acquired and examined for opinions or attitudes. Because it contains useful information, this stack of text data in Twitter is highly valuable. Data mining must be performed using specific methods in order to uncover this information. Text mining methods that can be used with the Natural Language Pre-processing method can be used to mine this data. Furthermore, the type of sentiment needs to be taken into account while deciding on the significance of mined data. By applying analytical feelings, this is accomplished. One popular social networking platform is Twitter. Twitter is used by users to share their Twitter with the public. There are now 330 million active Twitter users globally, and 18000 pieces of data are produced every second. The chirp that is provided may take the shape of news, commentary, arguments, or a variety of other sentences. As a result, Twitter has a lot of text with specific data. Typically, someone wants feedback from other people's viewpoints in order to make decisions. Direct inquiries might be made to obtain this view. It requires time and effort to meet people who are thought to inquire by directly approaching them. Twitter can be used as another resource for opinions. Tweets with many different opinions are offered by Twitter. The type of positive, negative, and neutral opinions must be used to distinguish this opinion. Additionally, these tweets were not categorised into the ones you were looking for. It remains common and important as a result.

II. Related Work

S. Symeonidis, D. Effrosynidis, and A. Arampatzis: In microblogging networks, sentiment analysis is a crucial tool for business and research purposes. We can draw insightful inferences about human behaviour thanks to machine learning methods that analyse human mood and comprehend human writing. The initial stage in text sentiment analysis is pre-processing, and employing the right methods with algorithms like Linear SVC, Bernoulli Naive Bayes, Logistic Regression, and Convolutional Neural Networks can increase the effectiveness of categorization. The detection accuracy is poor in this paper's lemmatization, number removal, and contraction replacement procedures.

J. Zhao and X. Gui: This study summarised the classification results of six pre-processing techniques using two feature models and four classifiers on five Twitter datasets. It also investigated the effects of text pre-processing techniques on the performance of sentiment classification in two different classification tasks. However, the author's usage of static Twitter data resulted in subpar training results.

X. Zhang, D.-D. Han, R. Yang, and Z. Zhang: In this research, the author describes the structure and dynamics of information dispersion in online social networks using empirical data that was scraped from Twitter. Propose a measurement with three metrics to indicate the efforts made by users to spread their knowledge, based on the unique methods for information retransmission on Twitter. It has been noted that while the majority of other users serve as middlemen when information is transmitted, a small number of users who excel at involvement can have a large impact. The loss of user activity and profile data will, however, come from the removal of missing data.

K. Schouten and F. Frasincar: This piece discussed The survey's summary of the most recent developments in aspect level sentiment analysis shows that the field has developed past its infancy.While in some instances a comprehensive strategy is described that can simultaneously do aspect detection and sentiment analysis, in others specific methods are offered for each of those two tasks. Given that language is a non-random, extremely complicated phenomenon for which there is a wealth of data, it is not unexpected that the majority of methodologies mentioned in this survey use machine learning to model language. But this paper introduces cutting-edge techniques for sentiment analysis.

S. Tsugawa and H. Ohsaki: They investigated the connection between a tweet's emotion and its virality in terms of diffusion volume and speed by looking at 4.1 million tweets on Twitter. As indicators of how popular a tweet was, they looked at the quantity of retweets and N-retweet time. They discovered that when the diffusion volume was high, negative tweets travelled more quickly than positive and neutral tweets and farther than negative tweets compared to positive and neutral tweets. However, the author focused on the connection between each tweet's sentiment and its virality. Calculating using the related feature approach is highly challenging.

S. M. Mohammad and S. Kiritchenko:In this paper, we compare the performance on a bag-of-words with that of multiple word- and character-based recurrent and convolutional neural networks. Additionally, we investigate whether a single model can be developed to predict all emotions using a common representation and whether the final hidden state representations can be applied to other categories of emotions. However, the author did work on bag of words strategies.

B. Plank and D. Hovy: Discourse parsing and sentiment analysis are two essential NLP tasks that are the subject of this researchthe development of three distinct recursive neural networks: one for sentiment prediction and two for the main discourse parsing sub-tasks of structure prediction and relation prediction. However, this labour is laborious and expensive because it is done by hand.

S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin: This study looked at how deep recurrent neural networks could be used to extract opinions at the phrase level. ESEs (expressive subjective expressions) are utterances that suggest feeling, emotion, etc. without directly transmitting them, while DSEs (direct subjective expressions) are explicit references to private states or verbal expressions of those states.
However, it is mostly resource and time expensive for the system.

J. Bollen, H. Mao, and X.-J. Zeng: The purpose of this study is to analyse electoral tweets for more subtly expressed information, such as tone (simple statement, sarcasm, hyperbole, etc.), sentiment (good or negative), emotion (joy, despair, wrath, etc.), purpose or intent (to point out a mistake, to support, to ridicule, etc.), and emotion.There are two sections: one on creating artificial classifiers to recognise these groups, and the other on marking text for sentiment, emotion, style, and qualities like intent.

S. M. Mohammad and P. D. Turney: Examine whether DJIA values and public mood, as determined by a sizable collection of tweets posted on twitter.com, are related or even predictable in this essay. The results show that, using relatively simple text processing techniques, changes in the public mood state may be seen from the content of enormous Twitter feeds, and that these changes exhibit highly variable responses to a range of socio-cultural reasons.The first negative, however, is the low barrier to account creation, and the second is the weak defence and poor reaction.

III. PROPOSED APPROACHES

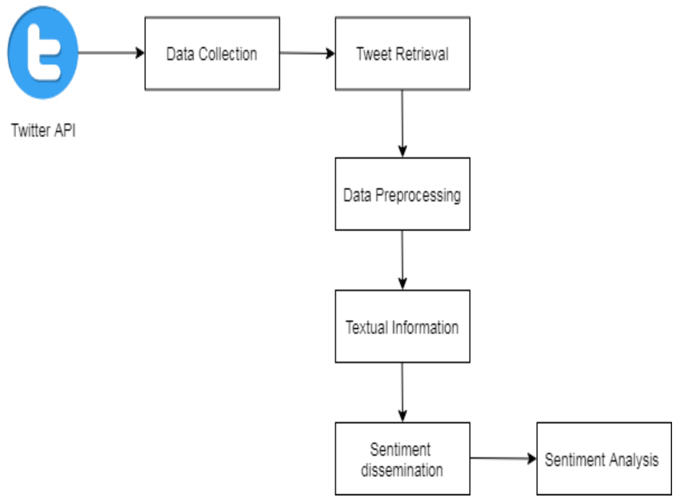

Fig 1. System Architecture

By looking at sentiment reversal, the situation where a tweet's sentiment and that of its rebroadcast differ, a theory of sentiment dispersion on Twitter was suggested. We examine the characteristics of sentiment reversals and suggest a model for predicting them.

We provide a machine learning technique to predict the sentiment polarity of each Twitter message that takes the interrelationships between the textual information of Twitter tweets and sentiment dissemination patterns into account. The likelihood that a tweet and its retweet will be correctly classified by a textual information based sentiment classifier will rise if the sentiment polarities predicted by that classifier are in line with the outcome of the sentiment reversal prediction. If not, the probability will decrease.. The textual content of Twitter messages can be blended with sentiment reversals in this way.

Algorithm:

1.Naïve Bayes

1. Given training dataset D, which consists of documents from several classes, such as Classes A and B.
2. Multiply the number of objects in class A by the total number of objects to determine the prior probability of class A.

   The prior probability of class B is equal to the number of objects in class B divided by the total number of objects.
3. Determine NI, the overall frequency of each class.

   Na is the overall frequency of class A events.

   Nb is the overall frequency of class B cases.

   4. Determine the conditional likelihood that a keyword will appear given a class:P (value 1/Class A) =count/ni (A)

   P (value 1/Class B) =count/ni (B)

   P (value 2/Class A) =count/ni (A)

   P (value 2/Class B) =count/ni (B)

   ……………………………………..

P (value n/Class B) =count/ni (B)

4. By using uniform distribution, avoid zero frequency issues.
5. Classify Document C based on the probability p(C/W)
   a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A)……. P(value n /Class A
   b.)Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B)……. P(value n /Class B)
6. Give the document to the class with the highest likelihood.

2. Random Forest

Random Forest is the algorithm employed in this case. The most well-known and effective machine learning algorithm is Random Forest.

Step 1: Pretend there are N training samples and M variables in the classifier.
Step 2: m input variables must be significantly smaller than M in order to determine the decision at each node of the tree.
Step 3: Select n times with replacement from all N available training samples before taking into account the training set. By predicting their classes, use the remaining cases to calculate the tree's error.
Step 4: Pick m variables at random for each node to use as the foundation for that node's decision. Based on these m training-set factors, determine the optimum split.
Step 5: No trees have been pruned (as might be done when building a typical tree classifier) and are all completely grown. A fresh sample is moved down the tree for predicting. In the terminal node it finishes up in, it is given the label of the training sample. The average vote of all the ensemble's trees is provided as the random forest prediction after this process is performed over each tree. i.e., the classifier with the most votes.

IV. RESULT AND DISCUSSION

The section displays the overall accuracy of the Random Forest and Naive Bayes algorithms. In comparison to the current method, this works delivers better sentiment analysis results.
The evaluation of the experimental results is noted as follows:
TP: True positive (number of instances that occurred as predicted)
FP: False positive (Inaccurately predicting the number of instances),
TN: True negative (predicted the number of instances as not necessary with accuracy)
FN: false negative(erroneously estimated the quantity of instances as not necessary),
We can determine four metrics based on this characteristic. Accuracy = TP+TN÷TP+FP+TN+FN
Precision = TP ÷TP+FP
Recall= TP÷TP+FN
F1-Measure = 2×Precision×Recall ÷Precision+ Recall.

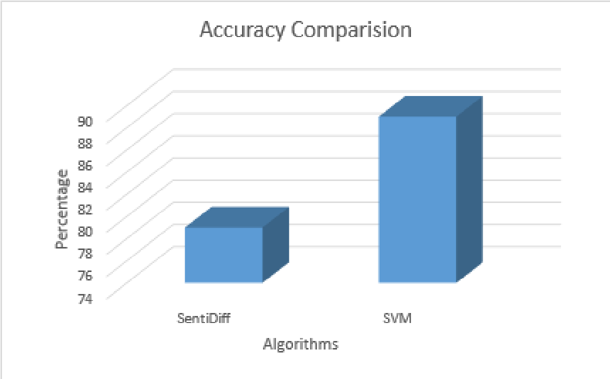

Fig. Comparison Graph

Sr. No.	**Existing System**	**Proposed System**
Precision	60.2%	65.4 %
Recall	85.5%	89.7%
Accuracy	Accuracy:72%	Accuracy:86%

Conclusion

The task of mining the polarity of the emotions portrayed in Twitter posts is significant and difficult. Due to the distinctive qualities of Twitter messages, the majority of existing Twitter sentiment analysis systems simply take into account the textual information of Twitter messages and are unable to deliver sufficient performance. Although recent research has demonstrated that patterns of feeling diffusion are closely related to the polarities of Twitter posts, current methodologies mostly rely on textual information from Twitter messages and disregard the transmission of information about sentiments. Take a first step towards fusing textual data and distributing emotions to improve Twitter's sentiment analysis, as inspired by recent work on the fusion of knowledge from many fields.

References

1. S. Symeonidis, D. Effrosynidis, and A. Arampatzis, “A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis,” Expert Systems with Applications, 2018
2. J. Zhao and X. Gui, “Comparison research on text pre-processing methods on twitter sentiment analysis,” IEEE Access, vol. 5, pp. 2870–2879, 2017.
3. X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, “Users participation and social influence during information spreading on twitter,” PloS one, vol. 12, no. 9, p. e0183290, 2017.
4. K. Schouten and F. Frasincar, “Survey on aspect-level sentiment analysis,” IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.
5. S. Tsugawa and H. Ohsaki, “Negative messages spread rapidly and widely on social media,” in Proceedings of the 2015 ACM on Conference on Online Social Networks. ACM, 2015, pp. 151–160.

6. S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets" Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015.
7. B. Plank and D. Hovy, Personality Traits on Twitter or How to Get 1,500 Personality Tests in a Week, Proc. of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 9298.
8. S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, Sentiment, emotion, purpose, and style in electoral tweets Information Processing and Management, vol. 51, no. 4, pp. 480499, 2015.
9. J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," J. of Computational Science, vol. 2, no. 1, pp. 1–8, 2011.
10. S. M. Mohammad and P. D. Turney, "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon," in Proc. of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. ACL, 2010, pp. 26–34.
11. R. S. Gaykar, V. Khanaa and S. D. Joshi, "Detection of Faulty Nodes in Distributed Environment using Machine Learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021, pp. 228-232, doi: 10.1109/ICAC3N53548.2021.9725478.
12. R. S. Gaykar, C. Nalini and S. D. Joshi, "Identification of Straggler Node in Distributed Environment Using Soft Computing Algorithms," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 1-5, doi: 10.1109/ESCI50559.2021.9396825.