IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Twitter Sentiment Analysis

[1] Ms.Sneha Kamble, [2] Mr.Dattatraya Dange,
[1] Assistant Professor , BIT Barshi,[2]] Assistant Professor , BIT Barshi

ABSTRACT

Social Media sites like twitter have billions of people share their opinions day by day as tweets. As tweet is characteristic short and basic way of human emotions. So, in this paper we focused on sentiment analysis of Twitter data. Most of Twitter's existing sentiment analysis solutions basically consider only the textual information of Twitter messages and strives to work well in the face of short and ambiguous Twitter messages. Recent studies show that patterns of spreading feelings on Twitter have close relationships with the polarities of Twitter messages. In this paper we focus on how to combine the textual information of Twitter messages and sentiment dissemination models to get a better performance of sentiment analysis in Twitter data. To this end, proposed system first analyzes the diffusion of feelings by studying a phenomenon called inversion of feelings and find some interesting properties of the reversal of feelings. Therefore, we consider the interrelations between the textual information of Twitter messages and the patterns of diffusion of feelings, and we propose an iterative algorithm called SentiDiff to predict the polarities of the feelings expressed in Twitter messages. As far as we know, this work is the first to use sentiment dissemination models to improve Twitter's sentiment analysis. Numerous experiments in the real-world dataset show that, compared to state-of-the-art text-based analysis algorithms.

Keywords—Text Mining, Machine Learning, Sentiment Analysis, Sentiment Diffusion, Twitter.

I. INTRODUCTION

Twitter, a popular micro blogging service around the world, has shaped and transformed the way people get information from the people or organizations that interest them. On Twitter, users can post status update messages, called tweets, to tell their followers what they are thinking, what they are doing or what is happening around them. In addition, users can interact with another user by replying or republishing their tweets. Since its creation in 2008, Twitter has become one of the largest online social media platforms in the world. Given the increasing amount of data available from Twitter, the polarity of the feelings of mining users expressed in Twitter messages has become a hot research topic due to its wide applications. For example, in analyzing the polarities of Twitter users on political parties and

candidates, different tools have been developed to provide strategies for political elections. Commercial companies also use Twitter sentiment analysis as a quick and effective way to monitor people's feelings about their products and brands.

This analysis is done by looking for opinions or sentiments from several sentences or tweets obtained. Therefore, this stack of text data in Twitter is quite valuable because it stores valuable information. To uncover this information, data mining needs to be done using certain techniques. Mining this data can be done using text mining techniques which can be combined also using the Natural Language Preprocessing approach. Furthermore, important data that has been mined needs to be determined by the type of sentiment. This is done by using analytical sentiments. Twitter is one type of social media that is often used. Users use Twitter to convey their Twitter to the general public. The number of Twitter users has reached 330 million people worldwide and every second produces 18000 data. The chirp delivered can be in the form of news, opinions, arguments, and several other types of sentences. This causes twitter to be rich in text that has certain data. In general, someone wants opinions from other people as input to determine decisions. This opinion can be done by asking directly. By asking directly, it takes time and effort to meet people who are believed to ask. Another way is to get opinions from Twitter. Opinions in the form of tweets provided by Twitter with a large amount. However, this opinion must be distinguished based on the type of positive, negative, and neutral opinions. In addition, these tweets have not been grouped according to the categories you want to find. So, it is still widespread and necessary. Sentiment Analysis is a technique widely used in text mining.

.

II. MATH

Sentiment analysis approaches In applying sentiment analysis, the key process is classifying extracted data into sentiment polarities such as positive, neu tral, and negative classes. A wide range of emotions can also be considered which is the focus of the emerging felds of afective computing and sentiment analysis (Cambria 2016). There are various ways to separate sentiments according to diferent research topics, for example in political debates, sentiments can be divided further into satisfed and angry (D'Andrea et al. 2015). Sentiment analysis with ambivalence handling can be incorporated to account for a fner-grained results and characterize emotions into such detailed catego ries such as anxiety, sadness, anger, excitement, and happi ness (Wang et al. 2015, 2020). Sentiment analysis is generally done to text data, although it can also be used to analyze data from devices that utilize audio- or audio-visual formats such as webcams to exam ine expression, body movement, or sounds known as multi modal sentiment analysis (Soleymani et al. 2017; Yang et al. 2022; Zhang et al. 2020). Multimodal sentiment analysis expands text-based analysis into something more complex that opens possibilities in the use of NLP for various pur poses. Advancement of NLP is also rapidly growing driven by various research, for example in neural network (Kim 2014; Ray and Chakrabarti 2022). An example would be the implementation of Neurosymbolic AI that combines deep learning and symbolic reasoning, which is thought to be a promising method in NLP for understanding reasonings (Sarker et al. 2021). This indicates the wide possibilities of the direction of NLP research. There are three main methods to detect and classify emo tions expressed in text, which are lexicon-based, machine learning-based approaches, and hybrid techniques. The lexicon-based approach uses the polarity of words, while the machine learning method sees texts as a classifcation problem and can be further divided into unsupervised, semi supervised, and supervised learning (Aqlan et al. 2019). Fig ure 1 shows the classifcation of methods that can be used for sentiment analysis, and in practical applications, machine learning methods and lexicon-based methods could be used in combination. When dealing with large text data such as those from Twitter, it is important to do the data pre-processing before starting the analysis.

This includes replacing upper-case let ters, removing useless words or links, expanding contrac tions, removing non-alphabetical characters or symbols, removing stop words, and removing duplicate datasets. Beyond the basic data cleaning, there is a further cleaning

process that should be implemented as well including tokeni zation, stemming, lemmatization, and Part of Speech (POS) tagging. Tokenization splits texts into smaller units and turns them into a list of tokens. This helps to make it convenient to calculate the frequency of each word in the text and ana lyze their sentiment polarity. Stemming and lemmatization replace words with their root word. For example, the word "feeling" and "felt" can be mapped to their stem word: "feel" using stemming. Lemmatization, on the other hand, uses the context of the words. This can reduce the dimensionality and complexity of a bag of words, which also improves the efciency of searching the word in the lexicon when apply ing the lexicon-based method. POS Tagging can automati cally tag the POS of words in the text, such as nouns, verbs, and adjectives, etc., which is useful for feature selection and extraction (Usop et al. 2017). 2.1 Lexicon-based approach The core idea of the lexicon-based method is to (1) split the sentences into a bag of words, then (2) compare them with the words in the sentiment polarity lexicon and their related semantic relations, and (3) calculate the polarity score of the whole text. These methods can efectively deter mine whether the sentiment of the text is positive, negative, or neutral (Zahoor and Rohilla 2020). The lexicon-based approach performs the task of tagging words with semantic orientation either using dictionary-based or corpus-based approaches. The former is simpler, and we can determine the polarity score of words or phrases in the text using a sentiment dictionary with opinion words. 2.1.1 Lexicon-based approaches with built-in library Examples of the most popular lexicon-based sentiment anal ysis models in Python are TextBlob and VADER. TextBlob is a Python library based on the Natural Language Toolkit (NLTK) that calculates the sentiment score for texts. An averaging technique is applied to each word to obtain the sentiment polarity scores for the entire text (Oyebode and Orji 2019). The words recorded in the TextBlob lexicon have their corresponding polarity score, subjectivity score, and intensity score. Additionally, there may be diferent records for the same word, so the sentiment score of the word is the average value of the polarity of all records containing them. The sentiment polarity scores produced are between [−1, 1], in which −1 refers to negative sentiment and +1 refers to positive sentiment. VADER (Valence Aware Dictionary and sEntiment Rea soner) is a lexicon and rule-based tool for sentiment analysis with a well-established sentiment lexicon (Hutto and Gilbert 2014). Compared to the TextBlob library, there are more corpora related to the language of social media, which may work better on a social media-type text that often contains non-formal language. From the results, the positive, nega tive, neutral, and compound values of tweets are presented, and the sentiment orientation is determined based on the compound score. There are several main steps of compound score calculation. Firstly, each word in the sentiment lexi con is given its corresponding scores of positive, negative, and neutral sentiments, ranging from −4 to 4 from the most "negative" to the most "positive." Heuristic rules are then applied when handling punctuation, capitalization, degree modifers, contrastive conjunctions, and negations, which boosts the compound score of a sentence. The scores of all words in the text are standardized to (−1, 1) using the for□mula below: where x represents the sum of Valence scores of sentiment words, and α is a normalization constant. The compound score is obtained by calculating the scores of all standard ized lexicons in the range of −1 (most negative) to 1 (most positive). The specifc classifcation criteria for both Text Blob and VADER are shown in Table 1. 2.1.2 Lexicon-based approach with SentiWordNet SentiWordNet is a lexical opinion resource that operates on the WordNet Database, which contains a set of lemmas with a synonymous interface called "synset" (Baccianella et al. 2010). Each synset corresponds to the positive and negative polarity scores. The value range of Pos(s) and Neg(s) is between 0 and 1. The process of SentiWordNet analysis is shown in Fig. 2. There are several steps in applying the SentiWordNet based approach. The frst steps are data pre-processing including applying basic data cleaning, tokenization, stem ming, and POS tagging. These steps can improve the time spent searching the words in the SentiWordNet database. For a given lemma that contains n meanings in the tweet, only the polarity score with the most common meaning is

considered (the frst one). The formula is as follows: We can count the positive and negative terms in each tweet and calculate their sentiment polarity scores (Guer ini et al. 2013). The sentiment score of each word or spe cifc term in the SentiWordNet lexicon can be calculated by applying Eq. (4): The SynsetScore then computes the absolute value of the maximum positive score and the maximum negative score of the word. For a term containing several synsets, the cal culation is as follows: PosScore = PosScore1 (2) NegScore = NegScore1 (3) SynsetScore = PosScore − NegScore (4)

TermScore = ∑k n=1 SynsetScore(r)⁄r ∑k n=1 1⁄r…..(5)

where n is a count number, the total score would be recorded as 0 if this term is not in SentiWordNet. The symbol k indi cates how many synsets are contained in this term, and if there are negations in front of this term, then, this sentiment value is reserved. Finally, we can add the sentiment scores of all terms to get the sentiment score of the tweets using the formula below:

SentiScore(s) = PosScore(s) + NegScore(s)

where p is a clean tweet with m positive terms and n negative terms. PosScore(p) is the fnal score of all the positive terms, while NegScore(p) represents the negative terms, and SentiScore(s) is the fnal sentiment score of tweets (Bonta et al. 2019).

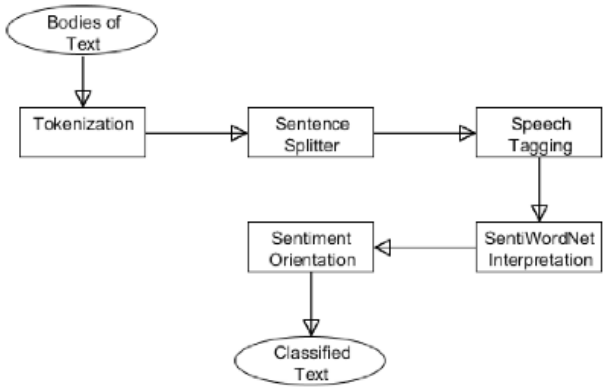

Fig 1: twitter analysis module

CONCLUSION

Polarity of mining sentiments expressed in Twitter messages is a significant and challenging task. Most existing Twitter sentiment analysis solutions consider only the textual information of Twitter messages and cannot achieve satisfactory performance due to the unique characteristics of Twitter messages. Although recent studies have shown that patterns of feeling diffusion are closely related to the polarities of Twitter messages, existing approaches are essentially based only on textual information from Twitter messages, but ignore the dissemination of information about feelings. Inspired by the recent work on the fusion of knowledge of multiple domains, take a first step towards combining textual information and spreading feelings to get a better performance of Twitter's sentiment analysis.

The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used. Sentiment analysis (SA) tells user whether the information about the product is satisfactory or not before they buy it. It helps them monitor their brand reputation, understand customer feedback, conduct market research, analyze competitors, and manage crises. With this tool, companies can make informed decisions and stay connected with their audience.

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. So we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance.

Sentiment analysis or opinion mining is a hot topic in machine learning. We are still far to detect the sentiments of s corpus of texts very accurately because of the complexity in the English language and even more if we consider other languages such as Chinese. In this project we tried to show the basic way of classifying tweets into positive or negative category using Naive Bayes as baseline and how language models are related to the Naive Bayes and can produce better results. We could further improve our classifier by trying to extract

more features from the tweets, trying different kinds of features, tuning the parameters of the naïve Bayes classifier, or trying another classifier all together.

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text preprocessing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

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 and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015.