Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, No.3, July : 2024

A SURVEY ON SENTIMENT ANALYSIS FOR TWEETS

Tapaswi Mestry Department of Information Technology Vidyalankar Institute of Technology Mumbai, India Tapaswi.mestry@vit.edu.in
Dr. Vipul Dalal Professor, Department of Information Technology Vidyalankar Institute of Technology Mumbai, India Vipul.dalal@vit.edu.in

Abstract—
Sentiment analysis deals with distinguishing and classifying opinions or sentiments expressed in force text. Social media is amassing a vast reservoir of knowledge wealth through tweets, status updates, blog posts, and more. Sentiment analysis of this user- generated knowledge is incredibly helpful in knowing the opinion of the gang. Twitter sentiment analysis is worrisome compared to general sentiment analysis due to the presence of dialect words and misspellings. Knowledge base approach and Machine knowledge approaches square measure the 2 styles used for assaying sentiments from the text. Public and private opinions on a wide range of subjects square measure expressed and unfold constantly via numerous social media. Twitter offers associations quick and effective thanks to assaying guests' views toward vital success within the business. This design cognitive content together with varied patterns for tweets along a side multiple strategies to discover the sentiment expressed in every tweet and if a tweet is real or not.
Keywords—NLP Sentiment analysis, machine learning, the influence of tweets, POS

I. INTRODUCTION

There are more than 100 million people who daily use Twitter and they tweet more than 500 million tweets every day. With a massive audience, Twitter has systematically attracted users to convey their opinions and perspectives regarding any issue, brand, company, or other interesting topic. According to this, Twitter is employed as an informative supply by several organizations, establishments, and firms. On Twitter, users are allowed to share their opinions in the form of tweets, using only 140 characters. This results in folks compacting their statements by exploiting slang, abbreviations, emoticons, short forms, etc. Along with this, people convey their opinions by using sarcasm and polysemy. Hence it’s even to term the Twitter language as unstructured. Sentiment analysis is employed to extract emotional context from tweets.

A lot of analysis has been done on Twitter knowledge to classify the tweets and analyze the results. In this project, we aim to predict the emotions from tweets by checking the polarity of tweets as positive, negative, or irrelevant. Sentiment analysis could be a method of explaining the sentiment of a specific statement or sentence. It’s a classification technique that derives opinions from tweets & formulates, Sentiment Classification is performed. Sentiments are subjective to the topic of interest. We area unit needed to formulate what reasonable options can decide for the sentiment it embodies. In the programming model, sentiment we refer to is a class of entities that the person performing sentiment analysis wants to find in the tweets. The dimension of the sentiment category is crucial to consider in deciding the potency of the model. For example, we can have a two-class tweet sentiment classification (positive and negative) or a three-class tweet sentiment classification (positive, negative, and irrelevant). Sentiment analysis approaches will be loosely categorized into 2 categories – lexicon-based mostly and machine learning-based. Lexicon's primarily based approach is unsupervised because it proposes to perform analysis exploitation lexicons and a rating methodology to gauge opinions. Whereas the machine learning approach involves the use of feature extraction and training the model using a feature set and some dataset.

We need to classify tweets into multiple types of sentiments by going further than just finding the polarity. We propose some classes such as sadness, happiness, and love to classify tweets according to their features extracted into these classes.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, No.3, July : 2024

II. LITERATURE SURVEY

Beakcheol Jang and Jungwon Yoon [2] conducted thorough measurements to understand the characteristics, both shared and distinct, of data from news sources and social media platforms. The identified variations are as follows: it's challenging to find consistent topics within both news and SNS. News typically covers official events, whereas SNSs are more responsive to personal interests. News tends to dwell on specific topics, while SNSs quickly transition between topics, with daily variations in the issues discussed. While news can pinpoint specific events with a single keyword, SNSs often require multiple keywords to locate relevant data.

Pros: Focuses on comparison between news and SNS, and provides insightful predictions. Cons: Requires a comprehensive keyword database.

Fang and Zhang [3] proposed a novel method for calculating polarities and strengths of Chinese sentiment phrases, employing probability values instead of fixed values for polarity strengths, unlike conventional methods.

Pros: Introduces a fresh approach using fuzzy logic. Cons: Limited applicability to the Chinese language.

Mondher and Tomoaki [4] presented a new approach for sentiment analysis, classifying sets of tweets into 7 different classes. Although achieving a modest accuracy of 60.2% for multi-class sentiment analysis, further optimization of training sets is deemed necessary for improved performance.

Pros: Supports multiple sentiment classes. Cons: Performance improvement required with better training data.

Aldo Hernández and Victor Sanchez [5] proposed a sentiment analysis method based on a linear regression model for tweets. The method effectively detects negative sentiments within specific contexts, aiming to predict responses of various stakeholders to hacking policies based on negative sentiments among Twitter users.

Pros: Effective in extracting opinions on specific issues using tailored datasets. Cons: Focus limited to specific issues, requiring dedicated datasets for prediction.

Manju Venugopalan and Deepa Gupta [6] developed a hybrid tweet sentiment classification model integrating domain-specific lexicons, unigrams, and tweet-specific features through machine learning techniques. Although enhancing classification accuracies, the model's focus on lexicons and unigrams outweighs sentiment analysis results.

Pros: Applies focused lexicons and machine learning techniques. Cons: Emphasizes techniques over sentiment outcomes.

Rincy Jose and Varghese S Chooralil [7] implemented a real-time, domain-independent Twitter sentiment analyzer utilizing sentiment lexicons like SentiWordNet and WordNet. By incorporating WSD and negation handling techniques, they achieved slight improvements in classification accuracy.

Pros: Utilizes online dictionaries for enriched datasets. Cons: Primarily addresses election sentiment prediction.

Anurag P. Jain and Mr. Vijay D. Katkar [8] proposed a system to classify tweets into different categories and verify their authenticity, employing knowledge base patterns, strategies, and machine learning approaches. By leveraging Twitter's API, the system enables real-time analysis, including identification of negative influences spread by users.

Mohd Fazil and Muhammad Abulaish [11] proposed a hybrid approach utilizing community-based features alongside metadata, content, and interaction-based features to detect automated spammers on Twitter. Exploiting spammers' interaction patterns, particularly low edge density among followers and followings, the approach promises effective spammer detection systems.

CONCLUSIONS

Sentiment analysis is used to solve a specific problem checking if a Twitter post is genuine or not. This method uses knowledge-based patterns, strategies, and machine-learning approaches. These methods are proposed to increase the accuracy of sentiment checks for tweets. Patterns can be used to evaluate if the tweets were influenced rumor or a genuine post by any user. By using the API of Twitter it is possible to work on live tweets than to work on offline data. Querying and fetching particular tweets from Twitter are possible by using its API. Identifying the spread of influence or negativity among users can prove valuable for a multitude of analytical endeavors.

VI. REFERENCES

[1] MondherBouazizi and TomoakiOhtsuki, “A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter”, pp. 2169-3536 on August 2017
[2]Beakcheol Jang and Jungwon Yoon "Characteristics Analysis of Data from News and Social Network Services", pp. 2169-3536 on March 2018
[3] Hai Tan & Jun Zhang, “Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness”, pp. 2169-3536 on April 2018
[4] MondherBouazizi and TomoakiOhtsuki, “A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter”, 2169-3536 on August 2017
[5] Aldo Hernández and Victor Sanchez published "Security Attack Prediction Based on User Sentiment Analysis of Twitter Data" in May 2016.
[6] Manju Venugopalan and Deepa Gupta explored "Sentiment Analysis on Twitter Data" in August 2015.
[7] Rincy Jose and Varghese S Chooralil worked on "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation" in November 2015.
[8] Anurag P. Jain and Mr. Vijay D. Katkar “Sentiments Analysis ofTwitter Data Using Data Mining”, pp.978-1-4673-7758-4 on Dec. 2015
[9] Gaurav D Rajurkar and Rajeshwari M Goudar presented "A Speedy Data Uploading Approach for Twitter Trend and Sentiment Analysis using HADOOP" in February 2015, as documented in pp. 978-1-4799-6892-3.
[10] Ahmed TalalSuliman, Khaled Al Kaabi, “Event Identification and Assertion from SocialMedia Using Auto-Extendable Knowledge Base”, pp. 2161-4407 on July 2016
[11] MohdFazil and Muhammad Abulaish, “A Hybrid Approach for Detecting Automated Spammers in Twitter”, on Nov. 2018.
[12] Yan Zhang and Weiling Chen, “Detecting Rumors on Online Social Networks Using Multi-layer Autoencoder”, on June 2017