

International Journal of
Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

This article can be downloaded from <http://www.ijerst.com/currentissue.php>

ATWITTER TREND MINING SYSTEM

KOLETI SAIPRASANNA¹, PASUPULETY JAAHNAVI², MUCHARLA RUCHITHA³

ABSTRACT:

Micro blogging is one of the main sources of trending news, events, and opinion in the modern day. One of them, Twitter, is often used to gather information on public comments and event updates. In order to find the tweets that most properly convey the trend's substance, this article will provide ways to filter tweets. It could also order the tweets based on their relevancy. The expected patterns in the users' natural language would serve as the basis for the rating process. A weighted score for each tweet in the sample space collected from Twitter is created using the mapping frequencies of word occurrence and associated hash tags for a certain trend.

Keywords: filter, trend, and mining for Twitter.

1.INTRODUCTION:

One of the greatest inventions of the "Internet" of the 21st century, where one person may interact with another globally with the use of a simple smart phone and internet, is made possible by the exponential expansion of computer science and technology. People used to connect with each other exclusively by email in the early days of the internet, and it was rife with spam. Back then, it was difficult to determine if an email was favorable or bad, or whether it was spam or not. As time went on, communication and data flow through the internet underwent a significant transformation, particularly following the advent of social media websites. The development of social media has made it very It's crucial to categorize the information into positive and negative words in order to stop any

kind of damage to society and to rein in people's antisocial conduct. Authorities have often detained persons in recent years as a result of their offensive and destructive social media posts[1]. For instance, a 28-year-old man was detained in Bengal for attacking Mamata Banerjee on Facebook, while a guy from Indonesia was detained for disparaging his country's police on Facebook. As a result, the situation is frightening, and it is essential to find such stuff before it is released since it is these bad contents that are making the internet a dangerous place and harming individuals. Let's say someone writes on social media, "Nonsense? Leave, you nerd. I said that,

1,2, 3 UG Student, Department of CSE, NARSIMHA REDDY ENGINEERING COLLEGE,
Maisammaguda, Kompally, Secunderabad, Telangana India. 500100

truth," it is clear that phrases like "Nonsense" and "Kiss off" are derogatory, making the statement poisonous. But in order to mine the toxicity technically, this remark must follow a certain process, after which a classification approach will be used to check the accuracy of the results. On the Kaggle.com data set, several machine learning techniques will be employed to categorize harmful comments. In order to address the issue of text categorization, this work uses six machine learning techniques: logistic regression, random forest, SVM classifier, naive bayes, decision tree, and KNN classification. Therefore, we will apply each of the six machine learning algorithms to the provided data set, compute their accuracy, log losses, and hamming losses, and compare them.

LITERATURE SURVEY:

1. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter.

AUTHORS: boyd, danah, Scott Golder, and Gilad Lotan.

Twitter is a microblogging platform that allows users to post messages ("tweets") of up to 140 characters. Participants use Twitter to communicate with individuals, groups, and the general public, so when conversations take place, they are frequently seen by audiences other than just the participants.

interlocutors. This essay investigates the act of retweeting as a method for engaging in "conversation." Retweeting has become commonplace on Twitter, but users do it in various ways and for different reasons. We emphasize the many ways in which authorship, attribution, and communicative authenticity are negotiated. This study illustrates the conversational practice of retweeting using a number of case studies and empirical data.

2 What individuals examine while examining Twitter: Identifying scholarly publications that mention Twitter

AUTHORS: Claire Warwick, Melissa Terras, and Irley Ann Williams Messages uploaded to the microblogging platform Twitter since its launch in 2006 have offered a rich dataset for

academics, inspiring the publishing of more than a thousand academic publications. This report seeks to find and categorize this published material in order to comprehend Twitter-based research. Design, technique, and approach The papers on Twitter were first located. Second, a taxonomy of the aspects of microblogging research was created after a survey of the literature. Thirdly, in order to examine method, topic, and approach, publications were qualitatively categorized using open coded content analysis based on the title and abstract. Findings The vast majority of works publishe Regarding Twitter, the focus is on the specifics of the messages delivered and the users. Several different methodological techniques are applied in the various indicated areas. Research limitations/implications This work examined the abstracts of all papers retrieved from databases using the term "Twitter." This has two major implications: first, the full papers are not taken into account, which means that works may be misclassified if their abstract is unclear; second, publications not indexed by the databases, such as book chapters, are not included. The research only looks at microblogging; other media are not taken into account when evaluating the method. Originality/value To far, there hasn't been a comprehensive study that examines the objectives and research methodologies of people who use Twitter. The main contribution of the study is to survey writings posted on Twitter up to the end of 2011. The taxonomy developed here will provide scholars researching Twitter-related subjects a framework within which to situate and root their work. Social networks are well-liked venues for friend connection, communication, and cooperation. Recently, researchers presented a new class of apps that make use of social network ties to boost efficiency and security in services like email, online surfing, and overlay routing. However, psychologists and sociologists have repeatedly questioned the practice of inferring meaningful relationships

only from social network connections, despite the fact that these programs often reference social network connectivity data to justify their designs. This raises the following query: Are social connections reliable measures of actual user interaction? If so, how can we quantify these elements to create a model for assessing socially enhanced apps that is more precise? In this essay, we answer that question by carefully examining user interactions on the Facebook social network. By measuring user interactions, we suggest using interaction graphs to provide significance to online social ties. We examine interaction graphs constructed from Facebook user traces and demonstrate that they have "small-world" characteristics at considerably lower levels than their social graph equivalents. As a consequence, the total network diameter grows noticeably and these graphs feature fewer "supernodes" with exceptionally high degree. We verify two well-known social-based apps using both kinds of graphs to measure the significance of our data (RE and SybilGuard). The findings provide fresh understanding of both systems and support our theory that social science

Applications should replace social graphs with actual evidence of user involvement.

Real-Time Twitter Trend Mining System, version 2.4

Min Song and Meen Chul Kim

By enabling users to access real-time online information channels free from time and place restrictions, social media is transforming how people now consume information. Additionally, it produces a vast quantity of data from which innovative information may be learned. Therefore, social media has made it very difficult for scientists to stay current with their field's progress. The majority of earlier research used broad methodologies, which often led to a lack of analysis. We present our real-time Twitter trend mining system, RT 2 M, which runs in real-time to analyse large stream datasets accessible on Twitter, in order to appropriately address these issues. The system provides features including retrieving term co-

occurrences, query-based visualization of Twitter users, user similarity computation, topic modeling to monitor changes in subject trends, and mention-based user network analysis. We also provide an empirical analysis of the 2012 presidential election in Korea. According to the case study, Twitter might be a helpful tool for identifying and forecasting the emergence and evolution of societal crises.

Networks may display several facets of user behavior.

5 Trend analysis and opinion mining using data from Twitter

Avneesh Jha, Ajay Singh Chahar, and Abhishek Singh Chauhan are the authors.

A great deal more data is now accessible online because to the surge in internet users throughout the world. People share their ideas and opinions on many facets of their life, including politics, entertainment, sports, economics, and more, through a variety of social media applications and online platforms. The goal of the "Trend Analysis and Opinion Minion" project is to collect all public views from Twitter and evaluate the most recent trends, which might be useful in deciding on marketing strategies, specific campaigns, and awareness-raising activities. This may be useful for detecting online instances of cyberbullying.

3. MECHANISMS

It is thought that the tweets don't provide much description on their own. A tweet on Twitter is limited to a maximum of 140 characters, which may include both alphanumeric and special characters. Due to this restriction, tweets must follow the telegraph style, where the importance of important phrases is given precedence above proper grammar. The range of the strategies suggested in this study extends to provide the tweets that were used as sample space scores so that they may be ranked as the most descriptive with the highest score. This may be used to familiarize a new user with trending material.

DEFINITION OF A PROBLEM

Some subjects will eventually become popular, while others won't. We want to foresee which

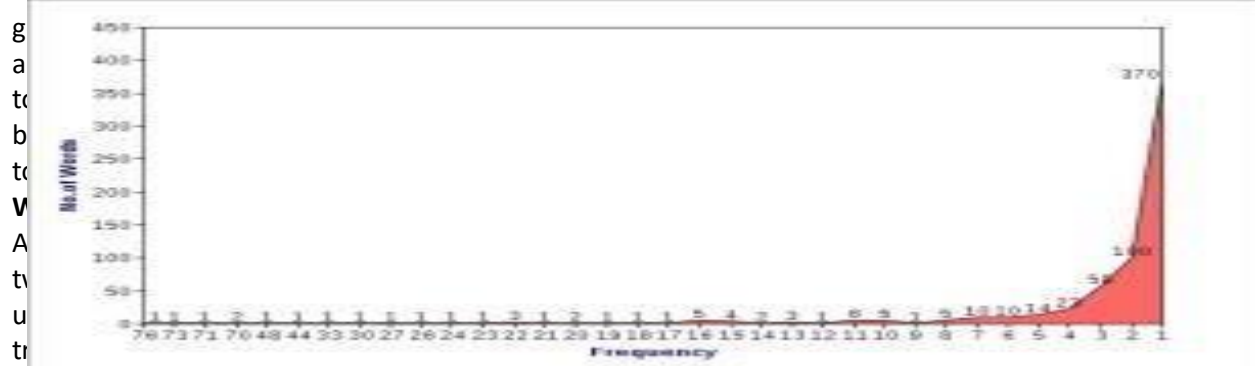
subjects will be popular. Apply algorithms to learn what the general population thinks about the subject to anticipate mood. Using trend analysis, you can forecast public opinion. It is significant because several researchers are developing methods for automatically extracting and analyzing the vast amounts of data from Twitter.

PURPOSE OF PROJECT

The most well-known social media platform that enables users to distribute and exchange information is Twitter. It keeps track of the user posts and finds the movement's most popular issue. These subjects are posted on the "Trending Topics" list. It demonstrates what is

popular tweet and other highly popular tweets. Tweets concerning the same thing or person often provide important information and may be expected to have more pertinent information. To update the previously created frequency table, the frequency data from tweets with the highest trending hashtag and the highest trending hashtag for the second time are collected.

After obtaining the frequency list, we rate the words to determine their weighted score. Each tweet's cumulative score is calculated using this weighted score, and it may be used to order tweets based on the importance of their content.



space's tweets, are ranked using a decreasing description index. Two dictionaries are also used by the algorithm. The first dictionary includes a list of terms, including as articles, prepositions, and conjunctions, that are more grammatical tools and have less bearing on the description of the content. All common nouns, adjectives, adverbs, verbs, and their derivatives are included in the second dictionary. Both will be referred to as "filter" and "cnfilter" from now on. The utilized sample space is saved in a file and divided by a character from the end of a tweet, such as „%%“. The frequency of each term used in the file containing the tweet sample space is discovered after the tweets have been collected. As a result, the "#" tags and the

„@“ tags. While determining the frequency, the URLs in the tweets are also disregarded. As a result, a list of terms and their associated frequencies is created and saved. It is now time to look for connections between the most

A tweet is only allowed to be 140 characters long, but when we run the suggested algorithms, we do not take into account whether the tweet is making use of all of the character space. The frequency of the target high frequency words should ideally be far higher than the frequency of the more common terms, whose contribution should be negligible, in order for them to contribute more to the score of a Tweet.

CONCLUSION

Given the outcomes of the algorithms developed for this research, it is possible to effectively explain the substance of the trends using Tweets. It is still not an effective approach to identify spam or RTs since the ranking algorithms logically hold true for the majority of tweet types. Algorithms for image and URL validation must be in place in order to implement those capabilities. Additionally,

these algorithms do not take into account the "favorite" characteristic linked to each tweet. A tally called a "favorite attribute" is increased each time a user upvotes a certain tweet.

Future Improvements

The length of the tweet is not taken into account in the present implementation. As per

REFERENCES

1. M. B. Rosson and D. Zhao. How and why individuals use Twitter: The function of microblogging in informal workplace communication. International Conference on Supporting Group Work, ACM 2009, Proceedings. ACM, 2009.
2. Gilad Lotan, Scott Golder, Boyd, and Danah. 2010. Twitter conversational aspects of retweeting are discussed in "Tweet, Tweet, Retweet." IEEE: Kauai, HI, January 6, HICSS-43
3. Claire Warwick, Melissa Terras, and Shirley Ann Williams (2013). Classifying academic publications that are relevant to Twitter: What people study when they research Twitter. Documentation Journal, 69 (3).
4. Twitter Search API. Twitter API documentation is available at <http://apiwiki.twitter.com>.
5. D. Sornette and R. Crane. Measurement of a social system's reaction function reveals robust dynamic classes. National Academy of Sciences Proceedings, 105(41), 15649-15653 (2008).
6. C. Faloutsos, J. Kleinberg, and J. Leskovec. Densification laws, decreasing diameters, and potential explanations in graphs over time. In Proceedings of the 11th International Conference on Knowledge Discovery in Data Mining, sponsored by ACM SIGKDD. ACM, 2005.
7. B. Y. Zhao, A. Sala, K. P. Puttaswamy, B. Boe, and C. Wilson. Social network user interactions and their effects. Fourth ACM European Conference on Computer Systems Proceedings. ACM, 2009.
- Why social networks vary from other forms of networks, by M. E. J. Newman and J. Park. Physical Review E, 68(3), 036122, September 2003.
9. Yang Q., Yu Y., Zhao K., Liu N.N., and Jin O. Using thematic information from

supplementary long texts to inform the clustering of short texts at the 2011 International Conference on Information and Knowledge Management