JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

COMPARISON OF DIFFERENT CLASSIFIERS FOR SENTIMENT ANALYSIS ON SOCIAL MEDIA DATA

A.Sairam[1], Dr.Amudha[2]

[1,2]Electronics And communication Engineering, SIMATS, Chennai-602105

[1]sairamfreeze@gmail.com, [2]ammukitcha@gmail.com

ABSTRACT

With the improvement of web technology and its evolution, for internet users there is a huge volume of data current in the web and a lot of data is generated too. The Internet has become an electronic learning forum for expressing thoughts and sharing views.This is where feature extractors and various machine learning classifiers are used. The extractors of the unigrams and features are unigrams, with definitive and negative keywords weighted. A system is designed that separates as a pair of components the function extractors and classificatory.This has become realistic that consumers are actually starting to look at product reviews that are available online before they purchase them. And for many companies the online assumption determines their product's success or failure. Sentiment Analysis therefore plays an significant role in businesses.In order to advance their products and in turn their status and help in customer satisfaction, businesses also wish to extract sentiment from the connected reviews. Semantics: The Algorithms classify the comprehensive feeling of a tweet.It is possible to use a semantic position labeler which shows which noun is allied with the verb and therefore the classification occurs.

INTRODUCTION

Twitter is a micro-blogging platform where users opt-in with others to obtain and send out extremely brief content — or tweets —. Or, in terms of the layman, it is a means of exchanging ideas and opinions in 140 characters or less. As of July 2009, controls more than 41 million users within less than three years, and is increasing rapidly.Within the 140-character cap, Twitter users tweet about any subject and follow others to get their tweets. The purpose of this paper is to research Twitter's topological features and its power as a new medium for sharing information. We've crawled around Twitter and got 41:7 million user accounts, 1:47 billion social ties, 4; 262 trend topics and 106 million tweets.

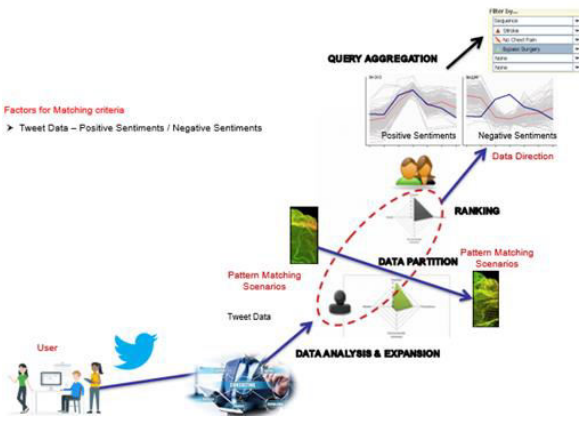

Figure:1

In its follow-up topology analysis , we find a non-power-law distribution of followers, a short effective diameter, and low reciprocity, all of which represent a divergence from known characteristics of human social networks[28]. We have rated users by the number of followers and by PageRank to classify influential on Twitter, and found two rankings to be identical.Ranking by retweets varies from the previous two rankings, suggesting an power discrepancy derived from the number of followers and that from one's tweets' visibility. We analyzed the tweets from top trending topics and commented on their temporal activity and user participation.We have categorized the trending topics based on the active time and the tweets which indicate that the majority of topics (more than 85 percent) are headline news or recurrent news in nature. A closer look at retweets shows that each retweeted tweet will hit an average of 1; 000 users irrespective of how many followers the original tweet is.When retweeted, a tweet is retweeted on next hops almost immediately, indicating the rapid distribution of information after the 1st retweet. To the best of our knowledge this work is the first systematic research on the entire Twittersphere and the distribution of information about it.Facebook users join or advise others. Unlike other online social networking sites , such as Facebook or MySpace, no reciprocation is expected in the relationship of following and being followed. Any other user will follow a user and the user being followed does not need to follow back. Having a Twitter follower means the user gets all the updates (called tweets) from the people that the user follows.Popular practice of responding to a tweet has evolved into a well-defined markup culture: RT stands for retweet, '@' followed by a user's address identifier

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

and '#' followed by a phrase represents a hashtag. Combined with a strict limit of 140 characters per posting, this well-defined markup vocabulary provides convenience for users with brief speech.The retweet mechanism empowers users to distribute information of their choice outside the reach of followers of the original tweet. Why do people get linked to Twitter? Who are the people with the most influences? Why is it people are worried about? Why does it diffuse knowledge through retweet? The goal of this work is to research Twitter's topological features and its power as a new medium for exchanging information.41:7 million user accounts, 1:47 billion social connections and 106 million tweets1 have been crawled. We start by analyzing the network and observing the distributions of followers and follow-ups, the relationship between followers and tweets, reciprocity, degrees of separation and hemophilia.First we rate users by the number of followers, the PageRank, the number of retweets and the quantitative contrast between them. The retweets ranking drives those with less than one million followers to top those with more than one million followers.Through our trending topic analysis we show which categories are grouped into trending topics, how long they last and how many users are involved. Finally, by retweet, we study the diffusion of information. We are building retweet trees and looking into their temporal and spatial characteristics.To the best of our knowledge this work is the first systematic research on the entire Twittersphere and the distribution of information about it. A user on Twitter holds a short profile of himself. The public profile contains the user's full name , location, a web page, a brief biography, and number of tweets. Often listed are the people who follow the user, and those who ignore the user.In addition, to crawl users not connected to the Twitter network's Giant Connected Module, we gathered profiles of those discussing trending topics in their tweets from June to August. The final user profile count we gather is 41:7 million.

LITERATURE SURVEY

The author(1) is attempting to use generalization and deletion techniques to protect the data. The data in the framework is evaluated for generalization, such as replacing (or recoding) a value with values that are less precise, but are compatible semantically. Suppression doesn't provide the user with the info.It is possible to secure the data by using generalization and suppression techniques and to have consistent values semantically. The main issue with this paper is that, in the suppression technique, there is no clear explanation on how the data will be secured. Given that the data isn't related semantically.The technique won't be successful in that situation. Text clustering(2,3) demonstrates clearly the essence of the leakage and the possibilities for preventing leakage. This defines how the dealer should "smartly" send agents data to increase the chances of identifying a culpable agent.By attaching fake items to the distributed collection the dealer will easily locate the guilt agent. We will get a good picture of the leakage, and quickly identify the guilty agent. This paper did not explain the type of impact that would occur on the system while adding the fake objects. Touching or amending sensitive data is not a suggested step.Clustering(4) can be extended to any database relationship that has attributes such that changes in any of its properties do not impact the applications. The tuple attributes within a tuple, bit positions in an attribute, and specific bit values are all determined algorithmically, under the control of a private key known only to the data owner.The methodology used in this paper is versatile and the leakage dependent on the attributes is very simple to identify. (4) Didn't clarify how the schema and clustering information would be given to the other person. Not sure how the owner will identify the criticality of the changes to be made to the data.Tracing procedures take advantage of the known structure or transformation properties when present, but also function in the absence of such information. The data modifications can be done while data storage and structure are also specified to track the transformation by using mining concepts. This paper did not concentrate on the latest methods that would automatically solve this sort of problem.No clear description in this tool regarding the security component. Document Networks(6) can be extended to any database relationship that has attributes such that changes in any of its characteristics do not impact applications.The tuple attributes within a tuple, bit positions in an attribute, and specific bit values are all determined algorithmically, under the control of a private key known only to the data owner.The methodology used in this paper is versatile and the leakage dependent on the attributes is very simple to identify. Didn't explain how the other user could get the information about the schema and watermark. Not sure how the owner will identify the criticality of the changes to be made to the data.

ALGORITHM IMPLEMENTED:

The current algorithm B-TREE Algorithm includes, Tree data structure that holds data sorted and enables searches, sequential access, insertions, and deletions in logarithmic time. It was difficult to access a significant amount of data from the secondary memory.Some of the algorithms were implemented to make our search very fast, to access the necessary data from the secondary memory B-trees are more efficient and faster B-trees. The proposed algorithm Temporal Pattern

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Search Algorithm requires a Collection of arrays, where each array includes all events of the same kind, sorted by time stamps. Particular element in the pattern using a binary search for the correct arrays.This allows TPS to miss several unnecessary events in personal history, which are considerably more costly per object. One Proposed Algorithm Backtracking Algorithm involves "Partial Candidate Solution" and a fairly simple check to decide if a suitable solution can be completed. Goes through this search tree recursively, in depth-first order, from the root down. The algorithm tests whether a correct solution can be completed to c.

IMPLEMENTATION:

Implementation includes the simple method of collecting information about how to use the twitter API. Apps include

- Valid Twitter Account
- Twitter Password
- API Key
- API Secret Key

When the authentication has been applied the application can enter the live twitter. The functionality allows clear tweet access focusing of the cut. In this the app offers the option to pull from your Twitter stream the ten most recent tweets.

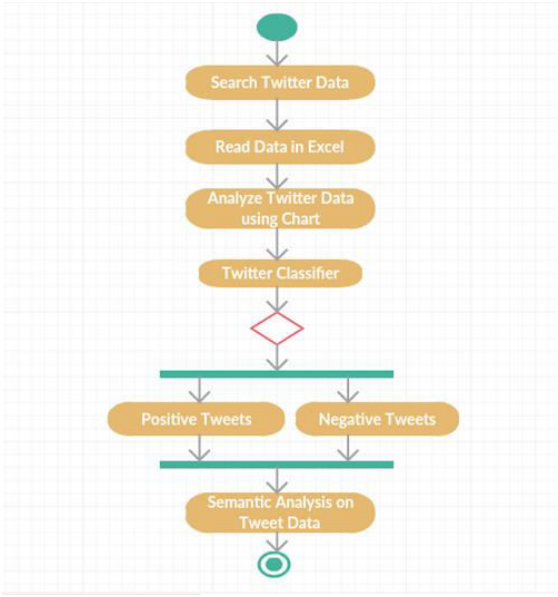

Figure:2

This can be done via the feature called as-function of the API object-home timeline).

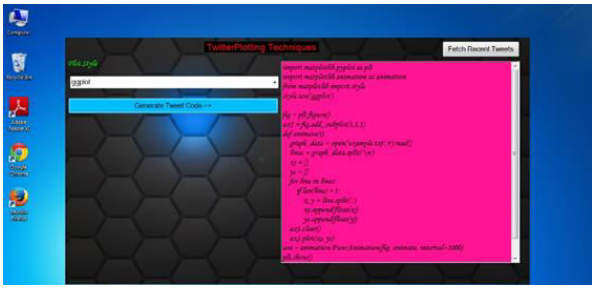

Figure:3

The data will then be processed after you have called the API. Twitter API result collection will be in JSON format. The content is taken out of the tweets via the tweet.text process. We need to look at the JSON provided by the Twitter API to access a particular attribute of each tweet message. The interface focuses on each tweet's "text" attribute and information about the tweeter.

The functionality extracts the latest tweets which contain a keyword that we are searching for. Monitoring a particular subject-If you want to monitor explicitly discussed issues in the Twitter community, this can be incredibly helpful to see if your company is being discussed.

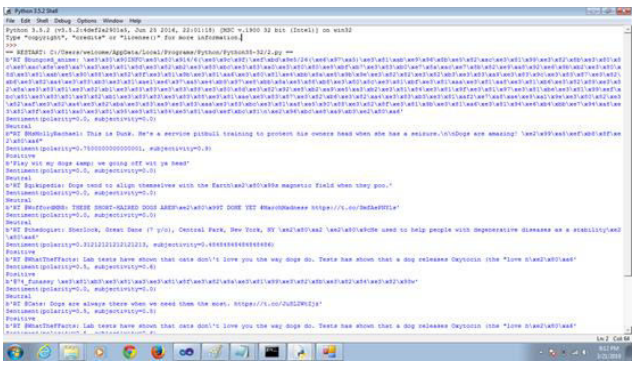

Figure:5

Run sentiment analysis on tweets to see if your company's aggregate view is positive or negative Generating a social graphic with the most popular users tweeting about your business or product.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

CONCLUSION

Growth in the field of opinion mining and sentiment analysis has been rapid and aims to explore the views or text present on different social media sites through machine-learning techniques with emotion, subjectivity analysis or polarity calculations. Despite the use of various machine-learning techniques and methods for analyzing emotions during elections, a state-of-the-art approach is urgently needed.To overcome these problems, this paper's contribution involves implementing a hybrid approach involving a sentiment Analyzer that incorporates machine learning.

FUTURE ENHANCEMENTS

Each application does have its own merits and demerits. Nearly all criteria were covered by the project. Additional specifications and changes can be easily made because the coding is of a standardized or modular nature in the key. Modification of existing modules or introduction of new modules will add improvements. More changes should be made to the framework, such that the website works in a way that is more appealing and useful than the present one.

We improve the current framework with the process of social network tagging with a tailored recommendation model based on the combination of user interest and tagging behaviors Potential development involves the web-based implementation of this method. With the tracking device, we agree to create the web application. To monitor the ship to reach the destination, where the ship is on motion. And we also plan in the future to update this web application to Android to make it easier for the consumer to access our website.

REFERENCES

1. Qian Wang, Cong Wang, Jin Li, Kui Ren and Wenjing Lou, 2009 - Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing

2. Charu C. Aggarwal, 07 January 2012, ChengXiangZhai, survey of text classification algorithms

3. Charu C. Aggarwal, Philip S. Yu, IBM T. J. Watson Research Center, Hawthorne, NY 10532, USA, University of Illinois at Chicago, Chicago, IL, USA - On Clustering Massive Text and Categorical Data Streams

4. Ralitsa Angelova, Stefan Siersdorfer - January 2006 - A Neighborhood-Based Approach for Clustering of Linked Document Collections

5. Inderjit S. Dhillon, Department of Computer Sciences, University of Texas, Austin, TX 78712, inderjit@cs.utexas.edu, Co-clustering documents and words using Bipartite

6. Jonathan Chang, David Blei ; Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, PMLR 5:81-88, 2009 - Relational Topic Models for Document Networks

7. Java Beginner's guide Book by Herbert Schildt

8. fThe Complete Reference Java 2 Fourth Edition Book by Herbert Schildt

9. Core Servlets and Java Server Pages Book by Marty Hall

10. Web Development with Java Server Pages Book by Duane K.Fields and Mark A.Kolb

11. My SQL Database Book by Marc Delisle

12. N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic,

and N. Bhamidipati. Hate Speech Detection with Comment

Embeddings. In WWW, pages 29–30, 2015.

13. A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov. Bag of

Tricks for Efficient Text Classification. arXiv preprint

arXiv:1607.01759, 2016.

14. Y. Kim. Convolutional Neural Networks for Sentence

Classification. In EMNLP, pages 1746–1751, 2014.

15. C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang.

Abusive Language Detection in Online User Content. In WWW,

pages 145–153, 2016.

16. J. Pennington, R. Socher, and C. D. Manning. GloVe: Global

Vectors for Word Representation. In EMNLP, volume 14, pages

1532–43, 2014.

17. Z. Waseem and D. Hovy. Hateful Symbols or Hateful People?

Predictive Features for Hate Speech Detection on Twitter. In

NAACL-HLT, pages 88–93, 2016.

[18] Karthik S, Annapoorani V, Dineshkumar S. Recognition And Tracking Of Moving Object In Underwtaer Sonar Images. International Journal of MC Square Scientific Research. 2016 Dec 12;8(1):93-8.

WEBSITE REFERENCE

1.Donation page: https://rcg.org>donate

2.Organ donation: https://en.m.wikipedia.org

3. Social media and organ donation Research paper, Authors: alexmorgan ,kushalpatel