

A Study on Traffic Analyses in Social Media Network of Twitter through Data Mining Techniques

M.C.S. Geetha, K. Divya Sri

Abstract - Social media helps to share the data around the globe. In this world, data is shared everywhere with the advent of social media, so that people from one place can share the data and their comments to others. Social Media provides comfortable way for sharing one information out of their interests. It can be used to contact their friends as well as making more friends and also search person with same kind of ideas and interests. It is very useful for any disasters occurred. At the time of tragedies social media paves a way for finding them and assist them. However sharing of the information in the social media like twitter, the traffic problem might be occurred due to heavy traffic. The social media user wants a minimum traffic to transmit the data without gap. In order to overcome this problem, this paper helps to study the comparison of various data mining techniques for smooth transfer of data.

Keywords - Social Media Twitter, Data Mining technique, content sharing and traffic analyses.

I. INTRODUCTION

In recent days, social media acts as an important role for all around the world. Social media offer many kinds of data which helps each one to contact other people in public and share their data. Social media helps to share, converse and connect with other personality or with a large spectators. Many people use these Social networking websites on the Internet each day to connect with millions of people. Some of the Social networking websites on the Internet are LinkedIn, Facebook, Twitter and Google Plus. At present era, for each and every piece of data shared in these sites like Twitter or Facebook, wall post, sharing photo, status and video, the user who is uploading have to decide to which of these friends or groups have to be sent. Twitter plays a vital role in the society and the conventional media. Online social networks are the websites that helps the users to provide contact with the other Internet users. Social networks accumulate all the distant data which the user needs. The Social Media provides comfortable way for searching the events, such as finding the natural disasters, since it has enlarged importance from various research group of people, as it provides a really important source of real-time data. The relationship between people in a social network is interdisciplinary. While transferring the information from one end to other end in the network the traffic problem may occur, because different people communicate at different time from different places

Manuscript published on 30 November 2018.

*Correspondence Author(s)

Ms. M.C.S. Geetha, Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, India,

Ms. K. Divya Sri, MCA Student, Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, India,

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The data is not secured, sometimes the data may get lost or congestion problem may occur. The data leakage may also sometimes happen.

In this research work, a twitter is taken as a tool for the social media. In the Twitter social network, there are around 400 million members and it allows the users to post 140-character tweets (messages) to their network of followers. The defective length of a tweet is given, the URL shorteners have suddenly developed the de facto method to share the friends on Twitter. The Twitter, because of lot of listeners and the information reach, attracts the spammers. The spammers don't have flexibility of having 140 characters. It is categorized as a micro blogging service. Micro blogging may be a group of blogging that allow users to send temporary text updates or micro media like images or audio clips. Micro blogging services that are missing from Twitter represent Tumbler, Friend Feed, social networking service is suitable to substitute small blogging services, and thereby tend to examine Twitter during this study. An important characteristic that is shared between small blogging services is their period nature. However web log users update their blogs once in many days. The users use the Twitter to write tweets several times during a particular day.

Operators will understand the substitute users and do the infrequently what they are threatening about at present. The customers frequently come back and check the position what people think. Several essential instances can be represented in their period of nature. Take the example of the case of powerful earthquake in Haiti, many footages were spreading through Twitter. While propagating the data in the immediate server may not be overloaded with the data at all the time, so traffic and congestion problem might occur. Because of this traffic, the data may get lost, damage and data leakage may also be occurred. This paper offers a various techniques of data mining approaches to reply the traffic problem in the social network.

II. LITERATURE REVIEW

In this paper [5] a tag based admission power of data was developed and shared in the social media sites. A change was made in the constructs access-control approach from photo organization tags.

All the photos has been included with the help of an admission network in order to record them for their friends. The worker can select a suitable preference and right to use the data. There are several other vital restrictions in the design.

First and foremost, our outcomes are not satisfactory by the contender employed for the photos they offered. Then the limitations is the usage of machine which helps to control the accessing rules. The algorithm did not have any permission to the event and outcome of tags and there is no forthcoming policy the challenger planned when access control was classified. From this conclusion, a few set of laws come about to evident out of the ordinary or unsystematic to the supplier, potentially driving them the decision was taken for the tags in the policy which can be public or private.

This paper was established by keeping the privacy settings that is depending on the typical of social loop [6]. From this criteria, there is possibility of explanation regarding the web to preserve individual information. With the help of the algorithm, the social circle was examined that helped to relate and finally the outcome of social circles are dealt by classifying friends for the privacy policies of the nature. The social circles plays a vital role by separating the theme however it do not display the topic. Questioning sessions are made on the topic so that they get inspired in order to share their personal information. The graphical representation was made after the users gives the reply.

This paper [7] offers method for the categorization of the image which is private and Search and in automatically sign the images which are private, and to make potential privacy-oriented image search. It unites the text Meta data images with that of graphic features to facilitate the security. In this method, the image features which are preferred such as histograms, edges and color help out to categorize between ordinary and artificial objects like the EDCV feature helps to choose the nonexistence of scrupulous objects such as SIFT. It is using the various classification models on a large scale dataset with the inaccessibility assignments that are completed with the social game.

In this paper [8], the system called an Adaptive Privacy Policy Prediction was developed and it has a free privacy mechanism for personalized policies. This system forces the user to upload the images based on the individuality and satisfaction of the metadata. It consists of two components: Adaptive Privacy Policy Prediction for Core and also the same system with Social. When a customer is uploading a picture, it will be instantly sent to the Adaptive Privacy Policy Prediction with Core. This core component arranges the picture and determines whether there is a need for demand for the other social component. The judgment policy being incorrect is the disadvantage because of not having knowledge of the Meta data information regarding the images. Both the Meta data and log data information are guided by the classification method.

In the paper [1,9] it is observed that there is lot of growth on Online Social Networks like Facebook, Twitter. These network intend to make communication socially to share the data and also there are a number of safety problems. Even though this network permit a single user to access their data, it need not pay compensation for any device to practice the privacy protection data with lot of users. Departure privacy violation basically helps to the dependable response of data and so at least single user are likely to keep private. The study was done for collecting the privacy and security issues in this network.

This network come across various types of problems such as fake identity, Sybil harassment and individuality clone attacks. The foremost aim of this social network is to supplement the privacy and security in the network which is one of the Quality of Service problem and so decline was done in the attacks. This paper affords a review on more thorough demonstration about the diverse attacks and privacy models regarding the growth of security and privacy [10].

Today, the social media is used more compared to previous generation to communicate lot of the personal information with others. With the help of this booming technology leads to isolating the rules by the users by distributing the large volumes of images to number of people. To give security for the information, automatic explanation of images are initiated, whose purposes is to create the meta data information about the images by using the novel method called Semantic interpret Markovian Semantic Indexing for retrieving the images [11]. The proposed system automatically understand the images by using hidden Markov model and features are extracted by using color histogram and Scale-invariant feature transform descriptor method. After understanding these images, semantic retrieval of images are made by using natural language Word Net for determining semantic assessment to interpret images in the database. Experimental outcome are available with higher retrieval performance when considering with the existing system.

Neha Mehta et al [12] analyze the web is the largest information repository experiential till date. Due to its huge size nevertheless, finding the significant information is not an effortless task. Different penetrating and web mining techniques are being employed by the current day search engine for the reason of information reposition from the web. Web document clustering is one potential technique to get better the effectiveness in information discovery process. The conventional web mining, method of web mining has complexity in managing confront posed by the gathering of data which is unclear and uncertain. Fuzzy clustering methods have the probable to manage such type of situations competently.

This paper summarizes the dissimilar characteristics of web data, the web mining basics and boundaries of existing web mining methods. The submission of use of Fuzzy logic with web mining is being converse with a view to highlight its significance in information retrieval. A relative study of dissimilar fuzzy clustering techniques with the predictable clustering technique has been discussed.

III.TRAFFIC ANALYSES USING VARIOUS TECHNIQUES

A.Geographical Distribution

Geographical distribution (GD) is one the method to automatically allocate the path for the data in the web. It routinely chooses the path for reassign the data in web. Distributing the nodes focuses more in the group

presentation of any real time data in the social media like Twitter, Facebook etc. Fault tolerance and preserving the competence of the system is handled by a well-distributed system. Geographical load balancing plays a role in separating a sequence of results from the available place and relocation of virtual machines computational jobs to geographically dispersed data centers in order to meet the service level agreements for the tasks and thereby reduce the operational cost of the web system. With the assist of this distribution, it chooses the path for scrutinize the traffic and it may choose the traffic-less path.

C.K-Nearest Neighbour (K-NN)

K-Nearest Neighbour algorithm is one of the effective one for judging the substitute path when the traffic happen. With this algorithm it contains the characteristic value for the nearer objects; with the help of the nearest object it routinely finds the path to relocate the data in the Social Media. This algorithm is used to find the nearest path in the web, comparison estimation of paths are close by location in the network to discover the traffic free path, this algorithm is helpful one to suggest to the traffic free path. This algorithm is a form of immediate based learning. The algorithm label analogous objects based on the neighboring feature space in the training set. The neighboring feature space may be resolute by determine the advance between the two feature vectors or by manipulative the Euclidean distance between the vectors. With the help of this algorithm it finds the nearest traffic-less path.

C. Ant Colony Optimization

Ants mainly used directly since, they together named an ant colony where they do all the important jobs as well as help to find the shortest distance travelled in order to search for food and to share the data with other ants. In the ant colony optimization, copies of shared intelligence of ants are part of the optimization techniques that helps to detect the uses of interacting with the computer. This procedure is constructed on the ordinary events done by the ant and the unexpected element such as pheromone is left by the ant which carries the search for the food and left out ant follows the pheromone.

The nearest route which has the highest pheromone is chosen since that will be the nearest route as the ant choose the shortest path which is said in the ant colony algorithm such that each ant will make sure that its record-set are protected and also the traffic which is entering may be going to the path that have the more possibility with highest pheromone. The ants also fully work to find the fresh food and simultaneously make use of the existing food sources return back to the nest. The main purpose of this method is to avoid the traffic between the nodes so that the ants will never turn back through a dead end by moving to the nodes for constructing a best solution set.

IV.COMPARISON ANALYSES OF VARIOUS TECHNIQUES

TECHNIQUE	ADVANTAGES	DISADVANTAGES
Geographical Distribution	It divert the path for low data only	Failed to balance the traffic
K-Nearest Neighbour	It divert the path to the nearest node when traffic occurred	Not known the nearest traffic path
Ant Colony	Traffic less, automatically finds the alternative path, Automatic finding the path the following data also keep going without distraction	High computation power

Table 1: Comparison Analyses

V.CONCLUSION

Nowadays Social networking is an important information sharing things. Social media offer many kinds of data which helps each one to contact other people in public and share their data. While sharing their content in the social media like twitter they may occurred the traffic problem for the social media web user. While transferring the data through twitter it may occurred a traffic problems. In order to overcome the problem, this paper provides the various data mining techniques. Here we have analyzed the traffic problem with the help of three algorithm. Based on our analyses ant colony provides a better result when compared to other algorithms.

REFERENCES

1. M. De Choudhury, Y.-R. Lin, H. Sundaram, K. S. Candan, L. Xie, and A. Kelliher. How does the data sampling strategy impact the discovery of information diffusion in social media? In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, pages 34{41, 2010.
2. A. Go, L. Huang, and R. Bhayani. Twitter sentiment classification using distant supervision. In CS224N Project Report, Stanford, 2009.
3. Harshita Rajwani, "Dynamic Traffic Analyzer Using Twitter"
4. Maximilian Walther and Michael Kaisser, "Geo-spatial Event Detection in the Twitter Stream", P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013.cS pringer VerlagBerlinHeidelberg 2013.
5. Peter F. Klemperer, Yuan Liang, Michelle L. Mazurek, "Tag, You Can See It! Using Tags for Access Control in Photo Sharing", Conference on Human Factors in Computing Systems, May 2012.
6. A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P. Tsang, "Social circles: Tackling privacy in social networks," in Proc. Symp. Sable Privacy Security, 2008.
7. Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova , I Know What You Did Last Summer!:Privacy-Aware Image Classification and Search , Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012
8. Anna Cinzia Squicciarini, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", IEEE Transactions On Knowledge And Data Engineering, vol. 27, no. 1, January 2015.

9. Ms.B.Hema and Ms.S.Sivagami, "Survey on secure and time confined image sharing on websites".
10. Anna C. Squicciarini, Mohamed Shehab, Federica Paci "Collective Privacy Management in Social Networks".
11. Sangeetha. J I, Kavitha, "An Improved Privacy Policy Inference over the Socially Shared Images with Automated Annotation Process".
12. Neha Mehta, Mamta Kathuria, Mahesh Singh, "Comparison of Conventional & Fuzzy Clustering Techniques: A Survey".
13. Vikram Singh and Balwinder Saini "An Effective Tokenization Algorithm for Information Retrieval System" CS & IT-CSCP 2014
14. Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo," Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", 2010.
15. H. Takemura and K. Tajima, "Tweet classification based on their lifetime duration," in *Proc. 21st ACM Int. CIKM, Shanghai, China, 2012*,pp. 2367-2370.
16. M.C.S.Geetha, I.Elizabeth Shanthi, "A Survey and Analysis on Regression Data Mining Techniques in Agriculture", *International Journal of pure and applied mathematics*, Volume 118 No. 8 Feb 2018, pg no. 341-347 ISSN: 1311-8080 (print) ISSN: 1314-3395 (online).
17. Parameswari, P., Manikantan, M.,(2017). GEO - INTELLIGENCE SYSTEM: A Framework for Agricultural Improvements, *International Journal of Pure and Applied Mathematics*, vol. 116, no.12, pp. 117-125.
18. Ratan Mishra and Anant Jaiswal, "Ant colony Optimization: A Solution of Load balancing in Cloud",in *International Journal of Web & Semantic Technology (IJWesT)*, Vol.3, No.2, pp. 33-50, 2012.