ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

A STUDY ON TRAFFIC ANALYSES IN SOCIAL MEDIA NETWORK OF TWITTER THROUGH DATA MINING TECHNIQUES

M.Bhuvaneswari[#1], K.MerlinJeba[*2], Dr.V.Srividhya[#3]

#M.Bhuvaneswari[1]
Research Scholar,
Department of Computer Science,
Avinashilingam University,
Coimbatore, India.
bhuvana.msc.mphil@gmail.com

#Dr.V.Srividhya[3]
Assistant Professor-SS
Department of Computer Science,
AvinashilingamUniversity,
Coimbatore, India.

***K.MerlinJeba[2]**
Research Scholar,
Department of Computer Science,
Avinashilingam University,
Coimbatore, India.
Merlin..mphil@gmail.com

Abstract- Social media provide a sharing of information across the world. Sharing the information is an important one in this real world, person on one area want to share their details, opinions, and comments to everyone or their friends. Social Media (SM) is a content sharing one which shares the user interests. Social networking can be used to keep in touch with friends, make new contacts and find people with similar interests and ideas. It also helpful for any disasters occurred. In the time of disasters it will helped the people lot to find them or to help them. While sharing their content in the social media like twitter they may occurred the traffic problem for the social media web user. User of the SM needs the traffic-less mechanism for continuous transfer of data. For this issue this paper provides a study of various techniques to overcome the issues.

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

I Introduction

Nowadays social media plays the significant role for everyone in this world. World provide a lot of information, messages, estimation data, etc. with these information in the people they desire to share their curiosity in public. It is one of the communication types for every user. Social media is the two way communiqué in Web 2.0 and it means to converse, share, and interrelate with an personality or with a large spectators. Social networking websites are the most renowned websites on the Internet and millions of people use them each day to employ and connect with other people. Twitter, Facebook, LinkedIn and Google Plus seems to be the most admired Social networking websites on the Internet. Today, for every single piece of content shared on sites like Twitter—every wall post, photo, status keep posted, and video—the up loader must make a decision which of his friends, group members, and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

other Twitter users should be able to access the satisfied. As a consequence, the subject of traffic on location like Twitter has conventional important concentration in both the research society and the conventional media. The expression "social media" refers to the wide variety of Internet-based and mobile services that permit users to contribute in online exchanges, supply user-created comfortable, or join online communities. Online social networks are websites that permit users to build connections and associations to additional Internet users. Social networks amass information distantly, slightly than on a user's personal computer. The use of Social Media for event uncovering, such as discovery of natural disasters, has gained a flourishing interest from research group of people as Social Media has become an enormously imperative source of real-time information. Social networking can be used to maintain in touch with friends, make new associates and find people with similar interests and ideas. The relation in a person's social network is multi-faceted. While transmitting the data in the network they may happens the traffic problem, since user of the social network shares the data at anytime and anywhere. While transferring the data in instantaneous the server may not be encumbered the data at all the time, so traffic problem, congestion problem may occurred. Because of this traffic issues the data may loss, damage and data leakage also happen. In our research we took a twitter as a tool in social media. Twitter, an established social network, has over 400 million members and it allows them to post 140-character tweets (messages) to their network of followers. Given the defective length of a tweet, URL shorteners have hurriedly become the de facto method to share links on Twitter. . Twitter, due to its large viewers and information reach, attracts spammers. The spammers have an incomplete flexibility with the 140-characters. Twitter is classified as a micro blogging service. Micro blogging may be a collection of blogging that authorize users to send temporary text updates or micromedia like images or audio clips. Microblogging services away from Twitter embody Tumbler, Friend Feed, social networking service is suitable to substitute small blogging services, and however we tend to particularly scrutinize Twitter during this study thanks to its superiority and knowledge volume. A significant characteristic that's common between small blogging services is their period nature. Though web log users usually update their blogs once each many days, Twitter users write tweets many times throughout a single day. Users will savvy substitute users do and infrequently what they're brooding about at present, users repeatedly come back to the position and check to ensure what people do. Many required instances represent their period nature: within the case of a very powerful earthquake in Haiti, several footages were broadcasting through Twitter. While broadcasting the data in instantaneous the server may not be encumbered the data at all the time, so traffic problem, congestion trouble may occurred. Because of this traffic subject the data may loss, damage and data seepage also occurred. This paper provides a various techniques of data mining approaches to answer the traffic problem in social network.

II LITERATURE REVIEW

Peter F. Klemperer [5] developed a tag based admission power of data shared in the social media sites. A move to that constructs access-control approach from photo organization tags. Every photo is included through an admission network for mapping the photo with the participant's friends. The contractor can choose a suitable partiality and right to use the data. Photo tags can be confidential as managerial or unrestrained based on the user needs. There are numerous significant restrictions to our study design. First, our outcomes are inadequate by the contestant we recruit and the photos they offered. A second set of restrictions apprehension our use of machine produce access-control rules. The algorithm has no permission to the circumstance and consequence of tags and no forthcoming into the policy the challenger proposed when classification for access control. As an conclusion, some rules happen to visible strange or random to the supplier, potentially pouring them in the way of unambiguous policy-based tags like "private" and "public.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Fabeah Adu-Oppong developed the privacy settings depends on the replica of social circles [6]. It make possible a web based clarification to preserve personal information. The practice named Social Circles Finder; mechanically create the friend's list. It is a method that studies the social circle of a person and classifies the attention of relationship and as a consequence social circles offer a important labeling of friends for environment privacy policies. The significance will distinguish the social circles of the theme but not show them to the subject. The subject will then be asked questions about their inspiration to share a piece of their individual information. Based on the act in response the function finds the visual graph of users.

Sergej Zerr proposes a approach Privacy-Aware Image categorization and Search [7] to mechanically notice private images, and to make potential privacy-oriented image search. It coalesce textual Meta data images with assortment of visual features to facilitate security strategy. In this the preferred image features (edges, faces, color histograms) which can help discriminate between natural and man-made objects/outlook (the EDCV feature) that can designate the continuation or absence of scrupulous objects (SIFT). It uses different classification models competent on a large scale dataset with isolation assignments accomplish through a social explanation game.

Anna Cinzia Squicciarini developed an Adaptive Privacy Policy Prediction (A3P) [8] system, a free privacy scenery system by mechanically produces modified policies. The A3P system levers user uploaded images based on the person's individual individuality and images satisfied and metadata. The A3P system consists of two components: A3P Core and A3P Social. When a user uploads a data like image, the image will be first sent to the A3P-core. The A3P-core organizes the image and determination whether there is a need to demand to the A3P-social. The drawback is mistaken confidentiality policy manufacture in case of the lack of Meta data information about the images. Also guide conception of Meta data log data information straight to inexact classification and also contravention privacy.

In the past years an incredible growth on Online Social Networks [1,9] like Facebook, Orkut and Twitter is seen. These OSNs not only proposition beautiful means for practical social communications and data sharing, but also elevate a number of protection issues. Although OSNs permit a single user to admittance to her or his data, they at the moment do not afford any device to put into practice privacy defense over data associated with large number of users, disappearance privacy infringement largely unreciprocated and leading to the credible acknowledgment of information that at least one user projected to keep private. They analyze the collection of privacy and security issues in OSN. OSNs come across diverse types of bother such a fake identity, Sybil harass, individuality clone attacks, The main aim is to supplement the privacy and security in OSNs which is one of the Quality of Service (QoS) issues and thus waning the attacks and harms. This paper is a survey which is more detailed to representation the different attacks and privacy models in OSNs with reverence to growth of security and privacy [10].

Usage of social media's amplified noticeably in today world which make possible the user to share out their personal information like images with the other. This improved technology leads to isolation disobedience where the users are distribution the large volumes of images crosswise supplementary number of peoples. To afford security for the information, perfunctory enlightenment of images are initiate which aims to create the meta data information about the images by using the novel method called Semantic interpret Markovian Semantic Indexing(SMSI) for reclaim the images [11]. The projected system automatically understand the images using hidden Markov model and features are extract by using color histogram and Scale-invariant feature transform (or SIFT) descriptor method. After understand these images, semantic repossession of images can be done by using Natural Language giving out tool specifically Word Net for determine semantic assessment of interpret images in the database. Experimental outcome make available enhanced recovery recital when appraise with the existing system.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

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 repossession 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. GD of the nodes subject a lot in the collective presentation of any real time data in the social media, particularly in case of the big scaled submission like Twitter, Facebook etc. A well-distributed system of nodes in web atmosphere is useful in handling fault tolerance and preserving the competence of the system. Geographical load balancing (GLB) can be distinct as a series of decisions about online obligation and/or migration of virtual machines (VMs) computational tasks to geographically distributed datacenters in order to meet the service level agreements (SLAs) or service deadlines for VMs/tasks and reduce the operational cost of the web system. With the assist of the GD it chooses the path for scrutinize the traffic and it may choose the traffic-less path.

b. K-Nearest Neighbour (K-NN)

K-Nearest Neighbour (*K-NN*) algorithm is one of the effective one for judgment 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 basically are straightforward being, they jointly forms a ant colony which do significant tasks including shortest path traversal to find food source and information distribution with other ants by producing pheromone. In the field of ant colony optimization, models of collective astuteness of ants are distorted into useful optimization techniques that discovery uses in computer networking. This algorithm is based on the natural activities of the ant in which a extraordinary substance called pheromone is laid down by the ant who goes out in search of the food and remaining ant follows the pheromone, The shortest path is selected in which greatest pheromone is there as that will be the shortest path as the ant choose the shortest path the ant colony algorithm each ant will preserve a record-set and the traffic that is incoming will be diverted to the path which has the more likelihood as in natural ant which select a path with highest pheromone. The ants work totally in search of new sources of food and concurrently use the existing food sources to shift the food back to the nest. The approach aims at competently avoid the traffic among the nodes and such that the ants never

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

come across a dead end for movements to nodes for building an optimum 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 occured	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. It can be used to keep in touch with friends, make new contacts and find people with similar interests and ideas. 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. For this issues this paper provide a various technique of data mining technique to solve the issues, we analyze the traffic problem with 3 algorithm, from our analyses ant colony provide 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 di_usion 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 classi_cation 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 1, 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] 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.