ORIGINAL RESEARCH PAPER

Computer Science

Volume - 7 | Issue - 2 | February - 2017 | ISSN - 2249-555X | IF : 3.919 | IC Value : 79.96

PREDICTING THE NETWORK TRAFFIC CREATING USERS FROM USER POSTING BEHAVIOR IN ONLINE SOCIAL NETWORKS

KEYWORDS: Social Networks, Data Mining, Traffic creators, Cat behavior and User behavior model.

Mrs.R. Gomathi

MCA, Research Scholar, Department of computer Science, Kovai Kalaimagal college of Arts and Science, Narasipuram, Coimbatore, TamilNadu.

Mrs.N. Vijayalakshmi

MCA., Mphil, Head of the Department,Department of Computer Science, Kovai Kalaimagal College of Arts and Science, Narasipuram, Coimbatore, TamilNadu.

ABSTRACT Social media acts a significant role among the media users. Social networks such as facebook, google+, foursquare, LinkedIn etc., have become extremely eminent over the world. OSN traffic is growing quickly and becoming significant, they want to learn the evolution of the traffic pattern of OSNs. In this paper, we propose an efficient swarm model to predict the traffic creation from user behavior model. In general, there prevail 12 types of social media users. Each user defines unique characteristics. The intention of this study is to discover the sorts of traffic creating social users. Inspired by the cat behavior, the user behavior model is designed into two modes, namely, i) Seeking mode and ii) Tracing mode. Firstly, the counts of cats are initialized. By estimating the optimal fitness value from support and confidence values, the users are placed to its relevant classes. By updating its position and velocity of each user, the traffic creating users from their spatial data is predicted. An experimental analysis shows the effectiveness of the systems.

I. INTRODUCTION

Online social networks plays vital role among the social media users. It is an interactive application which attracts several media users [1]. The season of mobile technologies and web-based technologies serves as communication medium among the social media users. The prevailed communication medium are the magazines, Internet forums, weblogs, social blogs, micro blogging, wikis, podcasts, photographs or pictures, video, rating and social bookmarking [2]. Most of the world is being surrounded by the social networks like facebook, skype, flicker, MySpace etc serves as better communication medium. This type of communication enables to mingle with person or group of individuals [3][4]. Most of the youngsters spend their time on the social media networks. Beyond the social communication, social networks are served as intermediate part in present years. The author in [5] depicted the merits and demerits of the social networks. The core benefit is the knowledge sharing among the group of individuals in concurrent time. This type of communication enables to improve their communication skills as well as the relationship.

The explorations of social networks have been employed in the business sectors. The business people make use of social networks to prove their brands in marketing field [6]. In the online environment, the brand promotion is done by the social networks. This assists them for arrangements maintenance the stronger position in the competitive world. This is collectively known as online advertising system [8]. Blog posts and tweets enable businesses to create communities, offer immediate feedback or assistance, and promote their products and services.

In this paper, the swarm systems used to solve the highly optimized problems. The communication between the swarm individuals operates on either direct or indirect. It is also known as population based algorithms. Nowadays, Social networks have been widely used as a source of information for the prediction of occasions from media like blogs, emails, Facebook, twitters etc. An occasion is defined as occurrence of the real entities that exist in particular time and space. We analyzed the location-based data i.e position and velocities to predict which type of social media users generating the traffic.

The paper is unionized as follows. Section I defines the basic terms and definitions in data mining and its importance of the study. Section II discusses related work. Section III confronts our proposed work. Section IV delivers the experimental analysis and lastly concluded in the Section V.

II. LITERATURE SURVEY

In this section, we presented the detection of events via social networks via two events namely, small-scale events and large-scale events. The small events depicts the analysis of traffic creating users, analysis of car crash etc whereas large scale events depicts the analysis of earthquakes, election etc. The deployment of small-scale events is now being established by several researchers. The author in [4] depicts the usage of social media networks and traffic creation in graphical manner. The case study, Japan was taken and their broadband analysis was done for the year 2004 and 2005. It was found that the outbound and inbound traffic analysis is higher for peak time 9pm to 11pm. The researcher also noted that 83% of the traffic created from TCP dynamic ports.

The author in [20] studied the analysis of access link saturation was done for 2006. The applications like BitTorrent, eDonkey, email and telnet was studied for the year 2006. The analysis was also made in the Chinese university campus network. The specified year 2005 was observed from 9pm until 10pm. It was concluded that outbound traffic are more than the inbound traffic. Many services provided for the students and employee in the Chinese university campus. In addition to, they also studied two applications MAZE and BitTorrent. Similarly, the authors in [18] informed that 58% of the HTTP traffic network and 25% of P2P applications network traffic. YouTube was studied for the network traffic measurements. Then the comparison was made for the wireless access networks and wired access networks and the researcher noted that wireless has less traffic than the wired networks. German wireless access network [11] connections of 250 households were studied for predicting the traffic creating users.

The author in [13] studied about the payload based classification for the period July, 2008. Most of the traffic generated from P2P, browsing and streaming networks. The sessions are recognized for each browsing categories. Signature based classification algorithm was deployed and reported that 37% of network traffic was from P2P systems and 25% of HTTP traffic [16].

And further, the study was extended on the streaming that determines 7% and online games contribute to 5%. The topmost applications are P2P and HTTP system which determines 50% of the network traffic. And it was proved on residential users from students and school employees of Greek school.

III. IMPROVED CAT SWARM OPTIMIZATION (ICSO) –PROPOSED ALGORITHM

In this section, we explain about the working model of an Improved Cat Swarm Optimization (ICSO). Some occasions in the OSN models

Volume - 7 | Issue - 2 | February - 2017 | ISSN - 2249-555X | IF : 3.919 | IC Value : 79.96

are not easily predictable. We intend to find optimal solutions for predicting the traffic creating events from the baseline of Association Rule Mining and Cat Swarm Optimization. The proposed architecture is given in fig. 1.

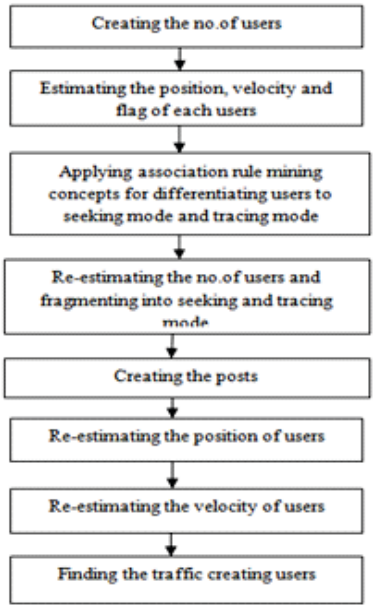

Fig.1. Proposed architecture

In general, the raw data contain user id, timestamp, location coordinates, retweet flag and text. Here, we took user id, timestamp, geographic coordinates and text of the tweet to discover which type of users creating the traffic. The objective of this study is to find out the traffic generating users by an Improved Cat Swarm Optimization (ICSO). The proposed algorithm works as follows:

i) Generating the counts of cats. i.e users.

ii) Every cat position is determined by its position, velocity and flag. The position is estimated from user-to-user association.

iii) Based on the present position, velocity of each user is discovered.

iv) Then the objective function is estimated by the eqn. (1).

$$MSD = [(x(t) - x_0)^2] \quad (1)$$

Where x (t) - Present Position of cats and x_0 – Cat's initial position.

v) The cat's behavior is observed into two modes, namely, seeking mode and tracing mode. Seeking mode describes the resting period of a cat whereas tracing mode describes the rapid movement towards target.

vi) Then, the support and confidence value for each user is defined for moving the cats to its relevant class i.e seeking mode or tracing mode. Support of the Cat system is defined as the percentage of posting between the users to the total number of social media users. Confidence of the cat system is defined as the percentage of posting that contain mutual link between the users to the total number of posts about the users.

vii) The seeking mode consists of five phases, namely, Seeking Memory Pool (SMP), Seeking Range of Selected Dimension (SRD), Self Position Consideration (SPC) and Counts of dimension to Change (CDC). SMP is used to define the size of seeking memory of each cat that suggests the points. The cat would pick a point from the memory pool, according to rules. SRD declares the variation in the selected dimensions. While in seeking mode, if a dimension is selected for variation, the difference between the old and new values may not be out of range, the range defines by the SRD. CDC tells how many of the dimensions will be varied. All these factors play important roles in seeking mode. SPC is a Boolean valued variable, and indicates whether the point at which the cat is already standing will be one of the candidate points to move. The SMP value is distincted from the SPC. By doing so, the position of the cat is updated at each iteration step.

viii) Tracing mode is used for finding the optimal solutions. The velocity for each dimension is given in eqn. (2)

$$v_{c1,d} = v_{c1,d} + r_1 * c_1 * (x_{pbest,d} - x_{c1,d}) \quad (2)$$

Where $X_{pbest,d}$ is the position of the cat C_1, who has the best fitness value. $x_{k,d}$ is the position of cat_k, c_1 is a constant and r_1 is a random value in the range of [0, 1]. Pbest is the best position at an iteration d. c_1 is an acceleration coefficient for extending the velocity of the cat to move in the solution space and usually is equal to 2.05. Again the cat's position is updated by the eqn (3):

$$x_{k,d} = x_{k,d} + v_{k,d} \quad (3)$$

The proposed algorithm covers the challenges like:

- Obtained best rules for achieving the target variables.
- Symmetry issue is achieved.
- Functional independency problems among the data in classes.

IV. EXPERIMENTAL RESULTS

In this section, we justified the proposed algorithm using Netbeans 8.1. We initialized some random users. By using our proposed algorithm, the network users are categorized into two modes, seeking mode and tracing mode. Atlast, relied upon the spatial data from pool of users, the traffic creating users are predicted.

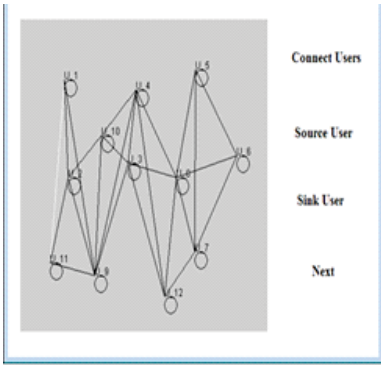

Fig.4.1. Connecting the nodes or users

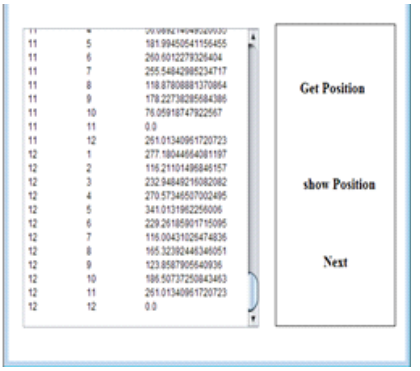

Fig.4.2. Displaying the position calculation with other nodes

Volume - 7 | Issue - 2 | February - 2017 | ISSN - 2249-555X | IF : 3.919 | IC Value : 79.96

Fig.4.3. Calculation of Velocity

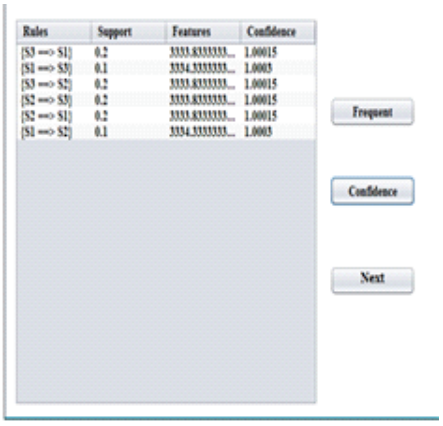

Fig.4.4 Applying ARM to find the users in seeking mode or tracing mode

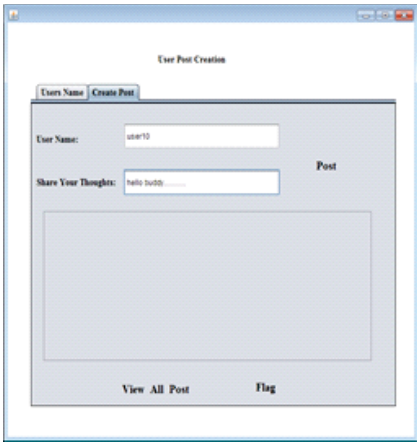

Fig.4.5 Creation of Post by the users

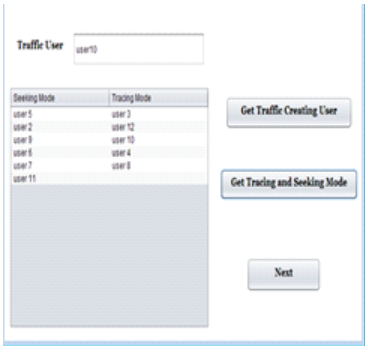

Fig.4.6. Displaying the users in seeking mode and tracing mode

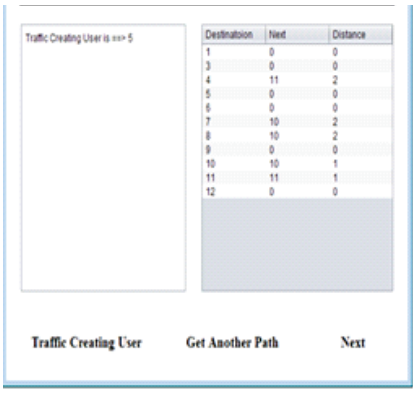

Fig.4.7. Predicting the traffic creating user.

V. CONCLUSION

In this paper, we propose an improved Cat swarm optimization (ICSO) that detects the traffic creating users from the predefined set of social media classes. Inspired by the swarm based approaches, the study in Online Social Networks (OSN) analysis is done. Using the spatial data of each user, the position and velocity rate of each user is determined. By estimating the support and confidence value of each user, the associations are formed. And thus, the traffic creating users are predicted. By our proposed algorithm, the challenges in OSN like dimensional reduction issue and symmetry issue are solved. An experimental result proves the efficacy of our proposed algorithm.

REFERENCES

1. R. Clegg, M. Withall, A. Moore, I. Phillips, D. Parish, M. Rio, R. Landa, H. Haddadi, K. Kyriakopoulos, J. Auge, R. Clayton, and D. Salmon, "Challenges in the capture and dissemination of measurements from high-speed networks," IET Communications, vol. 3, no. 6, pp. 957–966, 2009.
2. C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the sprint ip backbone," IEEE Network, vol. 17, no. 6, pp. 6–16, 2003.
3. K. Fukuda, K. Cho, and H.Esaki, "The impact of residential broadband traffic on japanese isp backbones," ACM SIGCOMM Computer Communications Review, vol. 35, no. 1, 2005.
4. K. Cho, K. Fukuda, H. Esaki, and A. Kato, "The impact and implications of the growth in residential user-to-user traffic," in Proc. of ACM SIGCOMM'06, 2006.
5. C. Kattirtzis, E. Varvarigos, K. Vlachos, G. Stathakopoulos, and M. Paraskevas, "Analyzing traffic across the Greek school network," in Proc. of the 14th IEEE Workshop on Local and Metropolitan Area Networks, 2005.
6. G. Xie, G. Zhang, J. Yang, Y. Min, V. Issarny, and A. Conte, "Survey on traffic of Metro Area Network with measurements on-line," in Managing Traffic Performance in Converged Networks, vol. 4516 of Lecture Notes in Computer Science, pp. 666–677, Springer-Verlag Berlin Heidelberg, 2007.
7. S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large net-works," IEEE/ACM Transactions on Networking, vol. 12, no. 2, pp. 219– 232, 2004.
8. G. Hasslinger, "ISP platforms under a heavy peer-to-peer workload," in Peer-to-Peer Systems and Applications, vol. 3485 of Lecture Notes in Computer Science, pp. 998–1010, Springer-Verlag Berlin Heidelberg, 2005.
9. T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos, "Is P2P dying or just hiding?," in Proc. of IEEE Globecom, 2004.
10. T. Karagiannis, A. Broido, M. Faloutsos, and K. Claffy, "Transport layer identification of P2P traffic," in Proc. of the 4th ACM Conference on Internet Measurement, pp. 121–134, 2004.
11. S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of P2P traffic using application signatures," in Proc. of the ACM 13th International Conference on World Wide Web, pp. 512–521, 2004.
12. P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: A view from the edge," in Proc. of ACM Internet Measurement Conference, 2007.
13. L. Guo, S. Chen, Z. Xiao, and X. Zhang, "Analysis of multimedia workloads with implications for internet streaming," in Proc. of ACM International World Wide Web Conference, 2005.
14. L. Guo, E. Tan, S. Chen, Z. Xiao, O. Spatscheck, and X. Zhang, "Delving into internet streaming media delivery: A quality and resource utilization perspective," in Proc. of the ACM Internet Measurement Conference, 2006.
15. E. Veloso, V. Almeida, W. Meira, and A. Bestravos, "A hierarchical characterization of a live streaming media workload," IEEE/ACM Trans-actions on Networking, vol. 14, no. 1, 2006.
16. G. Hasslinger, J. Mende, R. Geib, T. Beckhaus, and F. Hartleb, "Measurement and characteristics of aggregated traffic in broadband access networks," in Managing Traffic Performance in Converged Net-works, vol. 4516 of Lecture Notes in Computer Science, pp. 998–1010, Springer-Verlag Berlin Heidelberg, 2007.
17. J. Zhang, J. Yang, C. An, and J. Wang, "Traffic measurements and analysis of tunet," in Proc. of the 4th IEEE International Conference on Cyberworlds, 2005.
18. M. Ploumidis, M. Papadopouli, and T. Karagiannis, "Multi-level application-based traffic characterization in a large-scale wireless net-work," in Proc. of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, pp. 1–9, 2007.
19. H. M. Neto, L. Rocha, P. Guerra, J. Almeida, W. M. Jr., and V. Almeida, "A characterization of broadband user behavior and their e-business activities," ACM Performance Evaluation Review, vol. 32, no. 3, 2004.
20. M. Siekkinen, D. Collange, G. Urvoy-Keller, and E. Biersack, "Perfor-mance limitations of ADSL users: A case study," in Passive and Active Network Measurement, vol. 4427 of Lecture Notes in Computer Science, pp. 145–154, Springer-Verlag Berlin Heidelberg, 2007.