ANALYSIS OF SOCIAL NETWORKS

K.Anida Rashmi[1] and Dr.SS.Dhenakkaran[2]

[1]Research Scholar, [2]Professor

Department of Computer Science

Alagappa University, Tamilnadu, India

kanidarashmi@gmail.com

Abstract

Social Networks in the last decade has achieved incredible attention. This is qualified to the reason of accessing social network sites such as Face book, Hike, Instagram and other social network sites through the internet. Most of the people are becoming attracted in and depending on the Social Network for information and opinion of other users on various issue matters. It is important to convert opinion articulated by Social Networks users to useful information using data mining techniques. This underscores the importance of data mining techniques on Social Networks. Data mining techniques are capable of managing the three prevailing research issues with Social Networks data which are mass, noise and vitality. These techniques were used in information retrieval, statistical modeling and machine learning. These techniques use steps like data pre-processing, data analysis, and data interpretation processes. This paper reviews data mining techniques currently in use on analyzing Social Network and used k-means algorithm for clustering and fuzzy classification algorithm for classifying the data. And finally predicting which network is used by most of the people.

Keywords: Social Network, Social Network Analysis, Data Mining, K-means algorithm and fuzzy classification algorithm.

1. Introduction

Social network is a committed website allowing the user to communicate with each other by their post, videos, comments etc. Also they are web-based services that allow individuals creating public profile in a domain such that they can communicate with other users within that network. Social network has enhanced on the concept of technology of Web 2.0, by g the formation and exchange of User-Generated Content. Social network is a graph which includes of nodes and links representing the social relations on social network websites [1]. A node includes many entities and the relationships between them forming links. Social networks are important sources of online interactions and contents sharing, subjectivity [2], approaches, evaluation , influences, observations , feelings, opinions and sentiments expressions bear out in text, reviews, blogs, discussions, news, remarks, reactions, or some other documents. Before the advent of social network, the homepages was popularly used in the late 1990s which made it likely for average internet users to share information. However, the activities on social network in recent times seem to have transformed the World Wide Web (www) into its intended innovative creation.

Social network platforms enable rapid information exchange between users regardless of the location. Many organizations, individuals and even government of countries now pursue the

activities on social network. The network enables big organizations, celebrities, government official and government bodies to obtain knowledge on how their audience reacts to postings that concerns them out of the enormous data generated on social network. The network permits the effective collection of large-scale data which gives climb to major computational challenges. Yet, the application of efficient data mining techniques has made it possible for discover valuable, accurate and three dominant disputes with social useful knowledge from social network data. Data mining techniques are capable of handling network data viz., size, noise and dynamism.

The enormous nature of social network datasets requires information processing to be done automatically for analyzing within a reasonable time. Amusingly data mining techniques require huge data sets to mine the remarkable patterns from data, social networking sites appear to be perfect to mine with data mining tools. This form an enabling factor for advanced results for search in searching engines and also helps in better understanding of social data for research and organizational functions [3]. Data mining tools used for survey in this paper ranges from unsupervised, semi supervised to supervised learning. k-means is one of the simplest algorithm which uses unsupervised learning method to solve known clustering issues. It works really well with large datasets.

Cluster analysis is the task of partitioning data into subsets of objects according to their mutual "similarity," without using preexisting knowledge such as class labels vector support machine algorithm is a supervised machine learning algorithm which can be used for both classification or regression challenges. However it is mostly used in classification problems. The proposed method select the large dataset of the social networks like facebook, hike , instagram etc .First ,the dataset are preprocessed to eliminate noise data and the dataset is formatted using NP hard technique. Second the dataset is clustered using k-means algorithm to find the number of users in each network .Second, the dataset are classified using fuzzy classification algorithm to find the usage of networks in a particular time. And finally the peak hours of usage, the network most used by the users are predicted.

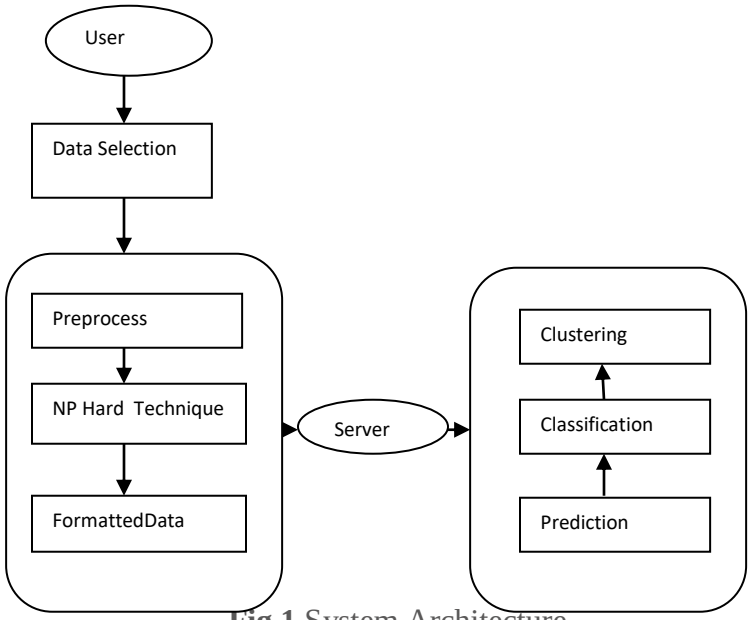

Fig 1. System Architecture

2. Proposed Method

The proposed method selects the large dataset of the social networks like Facebook, hike, Instagram etc. First, the dataset are clustered using k-means algorithm to find the number of users in each network. Second, the dataset are classified using fuzzy classification algorithm to find the usage of networks in a particular time. And finally the peak hours of usage, the network most used by the users are predicted.

2.1. Proposed Algorithm for Clustering

1. Clustering means grouping the particular objects here it's going to group the dataset according to the social media id, host and timing and segregate them in proper way.
2. Here clustering is done by k means algorithm , that means a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. In this paper the number of cluster is seven.
3. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result.
4. By using this algorithm we cluster the dataset that contains the attributes like social id, host, number of postings, login time, logout time etc.
5. Here the datasets are clustered into seven clusters of social networks namely facebook, hike, Instagram, Link-In, Mora, Twitter and watts-application.
6. According to this clustering the number of users and size of the data posted are recorded.
7. Using k-means algorithm, the datasets are clustered in minimum execution time.

2.2. Proposed Algorithm for classification

1. Any method for classifying data that allows attributes to apply to objects by membership values, so that an object may be considered a partial member of a class.
2. Here the fuzzy classification is used to classify the dataset according to timing (am/pm). And it would classify and given proper timing in which the user Sign in and Sign out the social media.
3. After the dataset is classified it can be easily stored in database and the values are fetched and used for further purpose such as to predict the peak hours of usage and to find the network mostly used by the users.
4. Using Fuzzy classification the peak timing result was found. The usage of social network in day time and time night are classified.

3. Experimental Results

Table.1. Result of k-means algorithm

S.No	Social Network	Count(Users)	Size
1.	Facebook	418	51234KB
2.	Hike	428	50766KB
3.	Instagram	435	53831KB
4.	Linked_In	478	58158KB
5.	Mora	449	53113KB
6.	Twitter	411	49904KB
7.	Watts_application	459	60459KB

Table.2. Result of Fuzzy Classification Algorithm

S.No	Social Network	Peak Hours a.m Session	Peak Hours p.m Session
1.	Watts_application	3264Hrs	1040Hrs
2.	Linked_In	3054Hrs	1110Hrs
3.	Hike	3015Hrs	996Hrs
4.	Instagram	2950Hrs	958Hrs
5.	Facebook	2848Hrs	1069Hrs
6.	Mora	2844Hrs	1150Hrs
7.	Twitter	2780Hrs	967Hrs

4. Conclusion and Future work

In this paper the outcome of the project is to predict the peak time and user count .For this purpose , we process all this steps i.e. in data preprocessing we are loading the data and we get the formatted data by using NP hard technique. And then next process is clustering it is processed by k means algorithm according to the social media Id, host etc. And the fuzzy classification is also done here to classify the dataset according to usage time. Then at last predict the peak hours and user count to find the network most used. The proposed algorithm is fast and efficient in terms of computational cost and works well even some assumption is broken. In future c means clustering algorithm can be used to cluster the result and it would predict the result in proper manner. And when conjugated with other enterprise system like SAP, direct sourcing of input data can also be implemented.

REFERENCES

[1] Junky Xuan; Xiangfeng Luo; Jibe Lu, "Mining Websites Preferences on Web Events in Big Data Environment," Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on , vol., no., pp.1043,1050, 3-5 Dec. 2013 doe: 10.1109/CSE.2013.152

[2] SanjeevPippa, LaishaBart, AK hila Krishna, Hana Gupta, Kuna Arora, Data mining in social networking sites: A social media mining approach to generate effective business strategies, International Journal of Innovations & Advancement in Computer Science (IJIACS), Vol. 3, Issue 2, April 2014.

[3] G Nandi and A Das, A survey on using data mining techniques for online network analysis, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.

[4] Osama Fayyad, Gregory Piatetsky-Shapiro, and Padraig Smyth. 1996. The KDD process for extracting useful knowledge from volumes of data. Common. ACM 39,11(November1996).

[5] RahulShoranReno, Gregory Mock, AshramKonner, Use of Big Data and Knowledge Discovery to Create Data Backbones for Decision Support Systems, Procardia Computer Science, Volume 20, 2013, Pages 446-453,ISSN1877-0509 .

[6] Chantal Yadav, Shulman Wang, Manor Kumar, Algorithm and approaches to handle large Data- A Survey, International Journal of computer science and network, vole 2, issue 3, 2013

[7] A K Jain, M N Marty, P. J. Flynn, 'Data Clustering : A Review', ACM COMPUTING SURVEYS, 1999

[8] Tina Zhang, Raghu Ramakrishna, and MyronLivy. 1996. BIRCH: an efficient data clustering method for very large databases. SIGMOD Rec. 25, 2 (June 1996),

[9] Hoer, P.; Hall, L.; Geldof, D., "A Cluster Ensemble Framework for Large Data sets," Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on , vol.4, no., pp.3342,3347, 8-11 Oct. 2006

[10] Jimmy Lin and DmitriRayon. 2013. Scaling big data mining infrastructure: the twitter experience. SIGKDD Explore. News. 14, 2 (April 2013), 6-19.

[11] FAHAD, A; Alcatraz, N.; Tara, Z.; Alarm, A; Y.Zomaya, A; Khalil, I; Froufrou, S.; Buras, A, "A Survey of Clustering Algorithms for Big Data: Taxonomy & Empirical Analysis," Emerging Topics in Computing, IEEE Transactions on, vol.PP, no.99, pp.1,