

# Prediction of Collective Behavior in Live Social Media

Kanchan U. Jadhav  
Dept. of Computer Engineering  
SKNCOE  
University of Pune

Nalini A. Mhetre  
Dept. of Computer Engineering  
SKNCOE  
University of Pune

## ABSTRACT

Human interest is the precious thing in the present world. Currently, online social networks such as Facebook, Twitter, Google+, LinkedIn, and Foursquare have become extremely popular all over the world and play a significant role in people's daily lives. Social media has provided simple communication and propagation of data, normally, through commenting, sharing and publishing. Now a day's social networking site performs more important activity for businesses or social work. If the interest of the user is already known then it can be easy target for any business or social activity. So it is important to find out behavior or interest of the users.

This work, intend to predict behavior of users in social media. Therefore our proposed system is going to show the behavior of the social networking user by extracting the online social networking data. Then construction of node graph and community graph to forms the grouping of similar behavior users. And then perform the clustering and classification for getting more accurate behavioral result of user.

## General Terms

Data mining, Behavior prediction,

## Keywords

Behavior Identification, K-means, Classification

## 1. INTRODUCTION

In current years, online social networks (OSNs) have spectacularly lengthened in all over the world. People access online social networking sites using both traditional desktop PCs and new emerging mobile devices.

With more than one billion people global, online social networking sites are a latest place of novelty with numerous challenging study problems. This paper, intend to provide a broad analysis of state-of-the-art scrutiny related to user activities in Social Networks from numerous viewpoints to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material.

In Blogs have rapidly evolve into a excess of social media platforms that enabled people to connect, share and discuss anything and everything, from personal experiences, to ideas, facts, events, music, videos, movies, and the list goes on forevermore. Social media has rapidly become a strong channel for news propagation Effortless publishing in social media has pushed people beyond "just" sharing and commenting the news they read to instantly report what is happening around them; think of local events that break out, e.g., political conflict, natural misfortune.

From a research perspective, social media can be seen as a microscope to study and understand human online behavior.

Behavior is a interest of persons in social networking surroundings. [3]. In this environment interest of one person is usually influenced by interest of other person or friend.

As per the information in October 2012, Facebook has 1.01 billion users accessing the site each month.

The speedy enhancement of Online Social Networks has fascinated a big number of researchers to investigate and examine this well-liked, ever-present, and great level service. I this paper, we focus on understanding user behavior in OSNs. Social Network user behavior cover up a variety of social actions that people can do online, such as sending friend request, topic sharing, browsing, chatting, and providing comments. Remarkably, these actions can be genuine or malevolent. Understanding OSN user behavior is important to different Internet entities in several aspects:

For Online Social Networking users, behavior study is essential to improve user skill and to know what is the exact interest of user. For example, there are several malevolent accounts in internet Social Networks. These accounts create unnecessary messages for valid users or does so many unwanted and unauthorized things. Therefore, recognizing and stopping unauthorized users are very essential to guarantee good user skill. Same way to know the interest of users is very important for so many businesses to enhance their business, such as advertising companies, political people etc.

This work contains different aspects for understanding user behavior in OSNs. First, perform the extraction of online social networking data and display two different graphs i.e. node graph and community graph. Based on these data this system perform the clustering of data by using K- means algorithm and then classification using SVM classifier.

The actual design of this paper is as given below. The section 2 describe the Motivation of System. Related work is explained in Section 3. Section 4 shows Proposed System Design, Section 5 explains Applications of the System. Section 6 shows Result of system. Section 7 & 8 gives conclusion and Future work. document.

## 2. MOTIVATION

Social networking sites are now part of daily life of population. This gives an opportunity to marketing or advertising companies.

In 2008 an article is published in New York Times, the article was presenting hurdles regarding the problem faced by advertisers on social networking sites.



Fig 1: Article in New York Times

As per the analysis all users on the social networking sites clicked on add but only 11% of those clicks lead to purchase. This was a very serious problem for advertising companies. Though they were able to reach to users they were not able to convert that opportunity in to money.

This gives rise for finding interest of users. Correct ads should reach to correct user and for this advertising companies want need and interest of users.

### 3. RELATED WORK

In early 2000 an analysis is done on available data for nearly one year for a particular website. Only text format data was available. With the available data and available technology only mathematical calculation for interest probability is calculated [5]. This work gave a solid base for further research.

Later the total dimensions of social networking sites are changed. Number of social networking sites is increasing with time. The data for analysis becomes huge and complicated for processing. Now the dataset available is complicated as no. of entities are attached with each other like group gathering similar hobbies sharing etc. The work done intend to mining out latent social dimensions [4]. The relational learning was done to find collective interest.

In the past few years enormous increase in number of users in social media happened. To deal with scalability edge centric clustering to extract sparse social dimension an innovation is stated [3].

A scalable K-means variant along with edge centric clustering is used to handle enormous data. Each edge is considered as one data instance and the associated nodes are the respective features. Along with the above technique a cosine similarity method to identify similarities between different node is installed[1][2].

### 4. SYSTEM DESIGN

Social networking sites are now part of daily life of population. The necessity of today's online social networking data is that the work should be done on online dynamic data. Fig 2. Shows no. of social networking sites which can lead to an enormous complicated database. Only dynamic extraction of dataset can resolve the problem of processing of such data.



Fig 2: Social Media

To performed this task a system with new architecture is designed. There are several modules which needs to be performed sequentially

#### 4.1 Extraction of online Database

A database is taken from social networking sites like Facebook .various facilities are provided by Facebook. By using one of application program interface (API) provided by Facebook the data is extracted from it. This API gives authority to user as a developer of application through which Login page application can be created.

So through this Login, data of users get extracted or fetched from facebook and it is possible to store this data in any application database.

#### 4.2 Node Graph Creation

The extracted data from social Networking site (i.e. Facebook) can be stored in any application database. This data is used for further application. A node graph is created based on friend relationship from the extracted data. Every user is considered as a node. A connectivity is shown if one user is friend of other and form an edge between both users. A friends node graph is generated by connecting several such edges by with each other. application database.

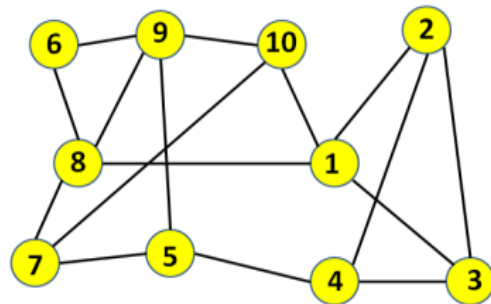


Fig 3: Node Graph

#### 4.3 Community Graph Creation

Likes of users are taken in to consideration. The community graph is created based on users likes. In a social networking site like Facebook users are having varieties of topics or application. Every user can have n number of communities because users can have n number of interest. Here a graph of users based on communities is shown. This graph shows the connectivity among the users who are having similar interest, so these similar interest users or nodes involved in one community.

Based on the community interests a community graph is created as shown in below figure 4.

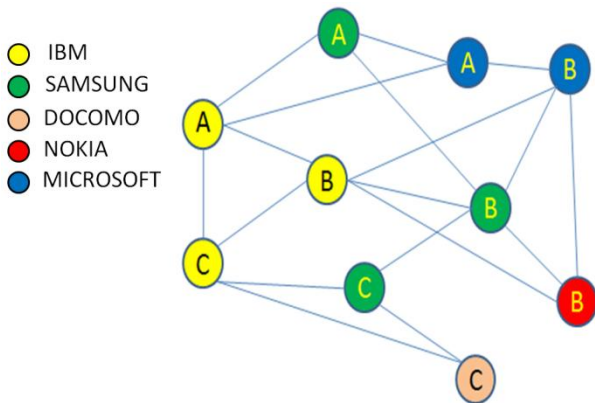


Fig 4: Community Graph

In above fig. there are five communities are shown that are IBM,SAMSUNG, DOCOMO, NOKIA and Microsoft and three users are A, B and C are shown. This graph shows which users are involved in witch community. User A,B and C shown with yellow color it means they are in IBM community. Green color nodes or users A,B,C are in SAMSUNG community. Same way 2 blue color nodes A and B are in Microsoft community. A single node C and node B are involved in DOCOMO and NOKIA community respectively. This fig shows the connectivity between all these same community members and connectivity between same members. Edge is form between all these users to create this community graph.

#### 4.4 Clustering by K-means Algorithm

The data fetched from Facebook is used for clustering. By using K-means algorithm clusters are formed from the data.

k-means clustering is a technique of vector quantization. k-means clustering seek to partition number of observations into k clusters in which each observation is a part of to the cluster with the nearest mean, serving as a prototype of the cluster. This fallout a separation of the data.

The dataset is divided into K clusters and the data points are randomly allocated to the clusters resulting in clusters that have approximately the same number of data points. For every data point it compute the distance from the data point to each cluster.

If the data point is nearest to its own cluster, keep it where it is. If the data point is not nearest to its own cluster, move it into the closest other cluster. Repeat the step until a complete pass through all the data points results in no data point moving from one cluster to another. At this position the clusters are firm and the clustering method ends.

#### 4.5 Use of SVM classifier for classification

After clustering SVM classifier is used to classify these data. This gives the correct prediction of the users behavior from these clusters.

In machine learning, support vector machines (SVM) are manage learning models with associated learning algorithms that evaluate data and distinguish sample, used for classification and regression. Given a set of training data, each put into one of two groups, an SVM training algorithm construct a model that allocate new examples or data into one group or the other, making it a non-probabilistic binary linear classifier. An SVM model is a illustration of the data examples as points in space, so that the examples of the

separate groups or class are separated by a clear gap that is as extensive as probable. New pattern are then linked into that same space and predicted to belong to a category based on which side of the gap they fall on.

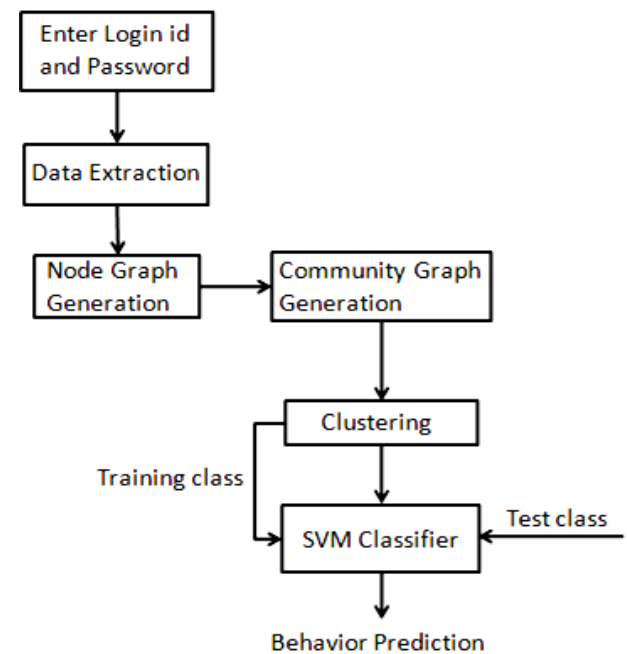


Fig 5: System Design

## 5. APPLICATIONS

**1. Advertising:** Now a day's User behavior prediction on a social networking site becomes a more important activity for advertising business.

A user whose behavior or interest is known can be a easy target for marketing. One can understand his needs and fulfilled his requirements by earning money.

**2. Search Engines:** The same thing for search engine. Search engine can provide more relevant result as per users requirement.

**3. Instant Messaging:** Instant messaging sites can provide a new chat partner who is having same interest. In many areas this behavior prediction can help to such sites to provide required friend.

**4. Election Campaign:** Election campaign finding voters interest in political party. Online interest is find out before actual election for deciding strategies of political parties.

**5. Professional Networking:** A professional networking sites like LinkedIn is now used by many organizations for finding people suitable to their job criteria. Directly from users profile they can find suitable candidates and some of his friends may also have same desired profile.

## 6. RESULT

The desired system is worked with the help of Facebook dataset. Facebook is a popular social networking site which is used by millions of users. Facebook user performs so many social activities such as friend requests, topic sharing, browsing, chatting, tagging and liking on these sites. It is observed that every user is having a distinct online behavior.

Based on above mentioned modules interest of a particular user is shown . The result will be based on the clusters in

which the user belongs to. users of similar interest are cluster together based on their likes from which behavior of the user whose interest is unknown can be find out.

## **7. CONCLUSION**

This paper is about online users behavior prediction on social networking sites. If interest of a group is known then interest of user belonging to that group can be find out. It is believed that persons having common interest are close friends of each other and only this rule a work for behavior prediction is based on.

By the method explains in this paper earlier a prediction of behavior of online social networking site user can be find out. This will help to advertising companies, Search engines and many more such agencies.

## **8. FUTURE SCOPE**

Social networking sites are improving day by day. These sites are providing more options to users. Users will perform many activities and which will generate more complex and complicated data.

To deal with such complex data a robust and efficient system design is required.

Also data size of such complex data will be high and to process this data time requirement will be more.

So a time efficient system design is required.

## **9. ACKNOWLEDGEMENT**

A special gratitude towards Prof. Nalini Mhetre for guiding and warm regards to Prof. Parikshit Mahalle for supporting and showing faith in author.

## **10. REFERENCES**

[1] Lei Tang, Xufei Wang, and Huan Liu, “ Scalable Learning of Collective Behavior,” volume 24, issue 6, IEEE, 2012.

- [2] L. Tang and H. Liu, ‘Toward predicting collective behavior via social dimension extraction,’ Intelligent Systems, volume 25, IEEE, 2010,
- [3] Lei Tang, Xufei Wang, and Huan Liu, Scalable Learning of Collective Behavior,” in CIKM ’09: Proceeding of the 18th ACM conference on Information and knowledge management. New York, NY, USA: ACM, 2009
- [4] L. Tang and H. Liu, “Relational learning via latent social dimensions,” Intelligent Systems,” in KDD 09, Proceeding of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, New York, 2009.
- [5] M. Newman, “Finding community structure in networks using the eigenvectors of matrices,” Physics Review E, Volume 74, No 3, 2006.
- [6] J. Neville and D. Jensen, “Leveraging relational autocorrelation with latent group models,” in MRDM’05: Proceeding of the 4th international workshop on Multi-relational mining, New York, 2005.
- [7] S. A. Macskassy and F. Provost, “Classification in networked data: A toolkit and a univariate case study,” J. Mach. Learn. Res., vol. 8, 2007
- [8] P. Singla and M. Richardson, “Yes, there is a correlation: - from social networks to personal behavior on the web,” in WWW ’08: Proceeding of the 17th international conference on World Wide Web. New York, NY, USA: ACM, 2008.
- [9] A. T. Fiore and J. S. Donath, “Homophily in online dating: when do you like someone like yourself?” in CHI ’05: CHI ’05 extended abstracts on Human factors in computing systems. New York, 2005.