

Group Prediction Identification of Hidden Relationship in Social Networks

Nithya.S¹, Dr.A.C.Kaladevi²

¹PG Scholar, Computer Science and Engineering, Sona College of Technology, salem, Tamilnadu, India

²Professor, M.E, Computer Science and Engineering, Sona College of Technology, salem, Tamilnadu, India.

Abstract—Social media is on the best way for registering our thoughts to friends and groups that produce voluminous data called big data. Those thoughts will reveal our emotions and the relationship with the particular person. Hence, getting group information or categories are important for advertisers. Advertisers deliver their product and related ads. The relationships in group or connection to some persons are difficult to find and media networks. Social Network Analysis predicts peoples' likes and dislikes in order to group that are in the form of hidden relationship. This work tries to use "Machine learning methods" like clustering and classifiers to extract the most sensitive characteristics from social them according to their preferences, hobbies, and education levels. These consequences help to understand the interactions between people, visualizing and navigating huge networks or supply basis analysis. For predicting customer behaviors for effective decision making. Though the proposed algorithm finds its application in almost all fields, we have chosen to advertise as the use case for best results.

Keywords- Big data, Social Media, Social Network Analysis, Clustering, Classifiers

I. INTRODUCTION

The increase in the number of social networking applications like Flickr, Face book evolved the way of communication between the people in the world. The amount of information passed during a single social interaction can vary from thousands to millions. As the forum concept of interaction keeps a thread of discussion alive for a large time period which increases the quantity and quality of the overall interaction. The data is processed appropriately; people can be differentiated and classified on the basis of their choices, ideas etc. This sudden growth was made possible due to the boom in the mobile technology and web-based applications along with it. With each user having mobile equipment, the data collected on the holistic view is very large. The users again vary from an individual to a private organization thus social networking covers all parts of the society and is not limited to the particular group. Apart from the communication point of view social networks provide

provisions for individual privacy. The social networks being open to all public equally the result of analysis on such information would be unbiased and would not be supporting any particular entity, which is the biggest advantage of this approach. The upcoming trend towards the pervasive computing is further changing the way of the collection of data from the social networks.

II. LITERATURE SURVEY

A. Machine Learning for Social Media Analytics

The blogs [13] over internet contain a lot of data, filtering the data about a particular product or brand is complex. The complexity is because of the fact that even similar product information is needed. This would help in finding the available market. This nature of analysis makes the keyword search method inefficient. And thus we explore additional techniques like text -based and network based techniques. While reading the blog contents, we segregate the relevant and irrelevant information. A standard text classification problem is formulated to filter relevance. Many graph based approaches exist to achieve the same.

B. Social Network Analysis Utilizing Big Data Technology

The data being generated by the mobile phones is increasing at a tremendous rate and the amount is huge already. This mobile generated data enables service providers to gather information about each user. The information collected from these users is used by the service providers for many purposes for example targeted marketing, selected offers etc. This information is quite vital for the service providers from their point of view and also to provide tailored service to each user. The work explores the advantages of the social media analysis [6] in telecom services and networks.

C. Social network analysis as a method in the data journalistic toolkit

In the democratic country like India, social [1] ill specifically corruption in public offices creates a lot of issues in the effectiveness of the country policies. Corruption erodes the overall market of the nation by the restricted and unaccounted money flow. Corruption thrives on discretion and thus is difficult to bring them to light. Techniques could be implemented in the public offices and public transactions to avoid any possibility of corruption. Monitoring agencies can be formed to guide the workflow of the public offices.

D. The Structure of Information Pathways in a Social Communication Network

Social media [12] are becoming the new hubs of the modern world where people from different part of countries and different age groups interact and share information. This nature of the social media is attracting the data scientists and researchers. The temporal dynamics of the exchange of information is studied by analyzing the online information, emails between the people participating over a certain time period. The temporal notion of distance is formed by calculating the time taken by certain information to move from one entity to other over a distributed computing system.

E. Modeling Blog Dynamics

The blogs network patterns [5] have been analyzed before and have resulted in an understanding that the pattern is bursty and mostly follows a certain law of power which express topological and temporal characteristics. But no prior methods have been proposed that can identify such patterns. There are models like a generative model that exploits naïve and intuitive methods for each blog and formulates a temporal characteristic set of the blogosphere. This approach also gives the topological network patterns like power laws for degree distributions etc.

III. PROPOSED METHOD

Marketers [2] often try to get the details of customers who are interested in a certain product type. Collecting this information is quite critical and beneficial to the business. Such data or information can be found from various social media networks. However, the social media frequently provides social network information of users. Although the relationships hidden in the relation between the events are complex and it cannot be shown from the social media. Hence, we are trying to use machine learning method to extract the likes and dislikes of people and group them according to their preference, hobbies and education level.

The figure 3.1 illustrates that the data set collected from the social networks. The KNN classifiers are applied to the group relevant problem statement. Later machine learning algorithm is applied within the groups to achieve the high rate of classification accuracy.

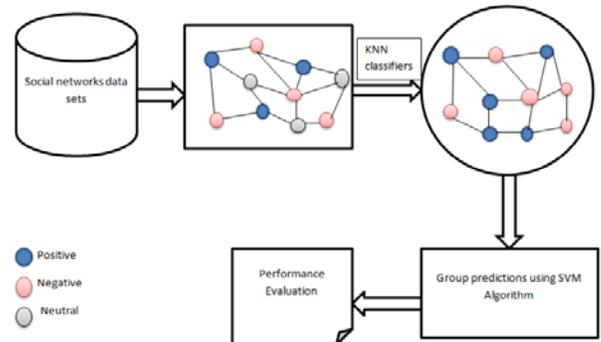


Figure 3.1: Block Diagram

IV. TECHNICAL SIGNIFICANCE

The key point is first to extract groups number from the social network, and then to use training data to get classifier and finally to use the classifier to predict remaining nodes in the network and which groups they belong to. Therefore, exactly getting the number of groups and choose fitting classifier are problems we need to deeply consider about. The approach is based upon machine learning; the advantage is that the algorithm is programmed to identify any new data set appearing apart from the known classes. The known patterns are identified and classified which is also the way any data mining concept would work. But the difference lies in the way each approach uses the information extracted. The machine learning approach[13] uses the information to update the knowledge base, subsequently improving the performance according to a machine learning algorithm for example the advertisements in the pages where any user browses changes according to the user choices and preferences.

V. PROBLEM FORMULATION AND TECHNICAL NOVELTY

The classification problem with social network data can be described as follows:

K categories $Y = \{1,2,3...k\}$;

A network $A = (V, E, Y)$ representing the interactions between nodes, where V is the vertex set, E is the edge set,

and each node V is associated with class labels Y whose value can be unknown; the known labels Y' for a subset of nodes V' in the network, where $V' \subseteq V$ [8].

Find: The unknown labels Y' for the remaining vertices.

For extracting groups from social network we are considering some method, might be spectral clustering [7] or others. The novelty is to use classifiers in machine learning to solve problem in social network.

VI. IMPLEMENTATION

Input: a social network A ;

Labels of some nodes in the network;

Number of social groups to extract;

Output: Labels of remaining nodes in the network

To use accuracy, precision or other method to evaluate the effectiveness of our methodology. The data set from the social network is collected and KNN classifier is applied to form the groups relevant to the problem statement, this is done to overcome the limitations of the machine learning classification technique. Later Machine learning algorithm is applied within the groups to achieve the high rate of classification accuracy. Here, the work of KNN classifier is to remove unwanted states to be passed to the machine learning algorithm so that the errors are not reflected in the knowledge base of the classifier. This key mechanism of the implementation is expected to give higher classification accuracy of the classifier.

VII. RESULTS AND DISCUSSION

The expected result focuses on getting information about unknown values. The unknown values are related to its relative surrounding values to some extent and mostly the values would correspond to those values. Only a small percent of such unknown value prediction would go wrong as expected. The known values are read and corresponded with the unknown values and its value can be predicted to a large extent by using machine learning techniques.

VIII. CONCLUSION

Social Network Analysis [1] also SNA maps and measures the relation and flows among people, groups, and other

interconnected entities. The analysis of such information along helps with pictorial and mathematical representations. Machine learning techniques are used to decide people on the basis of their choices, likes, and intellectual grades. The interactions between entities can be understood and represented as graphs and networks. This gives us a platform for generating data and information suitable for big data analysis to aid social network analysis. The machine learning has certain limitations due to its methodology to cater to new values and states. If an error arises and is not rectified the whole knowledge base is full of errors and the machine learning based classification accuracy would drop. To overcome this issue support vector machines can be utilized to check the error occurrences. This would keep the integrity of the machine learning based classification and the accuracy rate drop can be controlled. In future works instead of using only machine learning based classifiers other basic classifiers can be used to filter unwanted commonly occurring problems and these may be compared as well.

REFERENCES

- [1] Adriana Homolová "Social network analysis as a method in the data journalistic toolkit", Tilburg university.
- [2] Fortunato S, Barthelemy M Resolution limit in community detection. PNAS 104(1):36-41,2007.
- [3] Freeman, L.C. 'The Development of Social Network Analysis - with an Emphasis on Recent Events' in The SAGE Handbook of Social Network Analysis, pp.26-39, 2011.
- [4] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," KDD, 2008.
- [5] Goetz, J. Leskovec, M. Mcglohon, and C. Faloutsos, "Modeling Blog Dynamics," AAAI Conference on Weblogs and Social Media, 2009.
- [6] Jonathan Magnusson, "Social Network Analysis Utilizing Big Data Technology", 2012.
- [7] Luxburg, A tutorial on spectral clustering. Stat Comput 17(4):395-416, 2007.
- [8] Liu Y, Jin R, Yang L, Semi-supervised multi-label learning by constrained non-negative matrix factorization. In: AAAI, Orlando, FL, USA, 2006.
- [9] Lauchs, M. A., et al. 'Corrupt police networks : uncovering hidden relationship patterns, functions and roles' in Policing and Society, 21(1), pp. 110-127, 2011.

- [10] Marco Pennacchiotti, and Ana-Maria Popescu. "A Machine Learning Approach to Twitter User Classification", In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media ,2010.
- [11] Stray, J. 'Computational Journalism, Spring , Lecture 5: Social network Analysis', 2013.
- [12] Wasserman, S. & Faust, K. Social network analysis: Methods and Applications, Cambridge university press,1994.
- [13] Yan Liu, Pei-Yun Hsueh, Claudia Perlich "Machine Learning for Social Media Analytics", Yorktown Heights, NY 10598.