IJMTST | Volume: 2 | Issue: 05 | May 2016 | ISSN: 2455-3778

Alluding Communities in Social Networking Websites using Enhanced Quasi-clique Technique

Ms. Nidhi K.N.[1] | Mr. B.S. Yoga[2]

[1] Dept of CS&E, P.E.S College of Engineering, Mandya, Karnataka, India
[2] Dept of CS&E, P.E.S College of Engineering, Mandya, Karnataka, India

ABSTRACT

Social media is attracting global crowd rapidly. In websites such as Facebook, twitter etc one can share, view, like posts, such as images, videos, texts. Users also interact with each other. Communities are part of few such social networking websites. In a community people can learn more about their area of interest, share information on those topics, discuss about their perspectives etc. This paper recommends how community can be suggested to a user based on enhanced quasi clique technique.

KEYWORDS: *Community, User-interest, Social – networking, Text mining, Tokenization*

I. INTRODUCTION

Social networking has become a part of everyone's day to day life. It's a media that's building relationships, career, connecting people, educating them. Hence is in high demand regardless of one's age, sex, religion.

Community detection is one common issue's in social network mining. Community comprises of set of nodes with common interest put into a single group. An organization with the help of communities could find the customer specific interests based on the interaction. Thus, they could analyze few of their marketing strategies just by detecting communities. Such organizations thus may as well post various advertisements in specific communities benefitting them. Advertising based on specific interest is beneficiary not only to the organization, but also the members of the community. This is so because; community members are exposed to product of their interest. Thus the need for individual interest specific communities would be a major advantage.

Mining social networking websites for popular friends, strong group of friends and also community suggestion based on strong friends [5] are well known. Presently, community recommendation system is based upon strong friends and quasi clique technique [1]. That is, one classifies user's friends into strong friends, based on number of interaction with his/her friends.

Then, highest interacted friend is considered as the strong friend. Considering the communities of strong friend selected and using quasi clique approach, communities were suggested for the intended user. This paper, suggests an enriched community recommendation system using 'user area of interest' as an addition to the existing system. The implementation of the entire recommended system is done using text mining.

Text Mining

Data stored can be various forms, like image, picture, video, text etc. Growth in the division of text database is increasing in a rapid rate. When text data is attained from text databases (mining it using certain methodologies) for certain useful application then the mining is termed as text mining.

Text retrieval method falls into two major categories

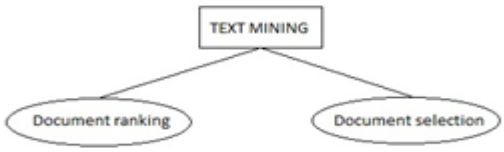

Fig1. Categories of Text Mining

Document Ranking

This methodology uses querying in order to rank all documents in order of relevance. Basic methodologies used here are mathematical foundation, algebra logic, probability etc.

In Boolean retrieval method user provides a link between various key words. This methodology is major use when the user has a good knowledge about the entire document.

Tokenization is method one such ranking methods where keywords are identified first. Then indexing is avoided for useless words by associating stop list with the document. Stop list comprises of irrelevant set of words called stop words. Some example of stop words are, as, if, the , or, with, for so on.

The proposed system in the paper makes use of tokenization.

Literature Survey

[1] In the paper "*Community Recommendation in Social Network Using Strong Friends and Quasi-Clique Approach*" the area of interest of the user's friends is checked and community is recommended. In this they never check if the user area of interest is similar to the strong friend's which is one drawback.

[2] "*Finding Popular Friends in Social Networks*" paper mentions how popular friends can be found using p-growth mining algorithm. Major advantage in the paper is time efficiency and space efficiency for sparse and dense datasets.

[3] "*Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques*", the sentiments are measured based on twitted data and tells if the statement is negative, positive or neutral. Thus sentimental analysis based mining cannot be implemented to predict the user area of interest based on their post.

[4] "*An Adaptive Approximation Algorithm for Community Detection in Social Network*", proposes an algorithm more precise than Eigen vector based algorithm based on modularity and computation time. Here the algorithm is used for complex and dense network as network contains many overlapping nodes and crossed edges. Based on the interest of interacting nodes common interest community is detected.

The further sections of the paper comprises of firstly, "Terminologies in need ", section which briefs about what are the major terminologies that are being used in the algorithm.

Secondly, the "Enhanced Quasi Clique Technique", section in which is the algorithm that's being used for the recommended system is being described. It comprises of two parts the first describing the how user area of interest is being identified and then how the quasi clique [1] technique is in a unique way merged with it.

The conclusion section, a brief about benefits of the proposed system is summarized.

Finally the references section, the research papers which were beneficially for this paper is being listed.

II. TERMINOLOGIES DESCRIPTION

Let A_i be set of user with F_i friends to whom, a community has to be suggested. Let F_{ai} be the the interaction strength between the user and his/her friends. C_{ai} be the set of communities in which users friends are present. The communities also have other users which are not in user's friend list, as it is a group where users of common interests are present.

A. Normalized Interaction Strength

We calculate the normalization value of interactive strength for every friend by dividing the interaction strength with the total number of interaction to his all friends.

B. Cumulative Normalized Interaction Strength

$$C_n = C_{n-1} + B_n$$

Let, B_n be the normalized strength of the next consecutive friend in the list of all his friends.

C. Minimum Strength

Minimum strength (min_{strn}), minimum strength value (min_{strn}) is the threshold or base value for recognizing strong and weak friends.

D. Strong Friends

Strong Friends (S_{ai}), we define S_{ai}= {s1, s2…sn} as a set in descending order based on ($nis_{ai,fi}$). When the cumulative normalized interaction

strength ($nis_{ai,fi}$) exceeds the minimum strength value (min_{strn}) we consider up to those friends as a strong friend set. The set indicates that the user (A_i) is most likely going to interact with these friends.

III. Enhanced Quasi Clique Technique

Algorithm Steps

A. Identification of User's Interests

a) The users posted text information in various forms such as, image description, video description or general text posts are to be collected from the server (database).

b) Tokenization of the extracted information, so as to attain useful keywords and separate them from the stop words.

c) The administrator is required to create certain community type and set certain key words under each community type to classify each community.

d) Comparing keywords set by the administrator and the extracted keywords of every post each post is classified into certain community type.

e) Then the community type with highest post is taken to be the user area of interest.

f) Therefore, identified the user's area of interest and thus must be recorded in database system.

B. Strong Friends Prediction

a) Firstly Friend database and community database are scanned.

b) Normalization [normalized interaction strength] is found for all the user's friends and then arranged in descending order.

c) The cumulative normalized interaction strength is predicted using the equation

   a. $C_n = C_{n-1} + B_n$

d) Here C_{n-1} is the cumulative normalized interaction strength of the previous and B_n is the normalized interaction strength of every user.

e) A certain Minimum strength (min_{strn}) is set.

f) The threshold is set so as to separate strong friends S_{ai} and weak friends W_{ai} of the user.

g) Consider a condition

   a. if ($C_n \geq min_{strn}$) is set.

h) If the above mentioned condition is true then, for the friend for which the above condition becomes true, till that friend, all the user's friends are considered as strong friends.

i) Next, we consider those communities (C_{ai})[based on user's area of interest] which contains strong friends.

j) We then calculate the normalized interaction strength (min_{strn}) based on the community.

k) Finally we display the communities (C_{ai}) in the descending order.

l) For those communities which user is interested in but doesn't contain any of users strong friends then we do the following steps:

   1. Calculate the normalized interaction strength(min_{strn}) for all communities
   2. We finally display the communities (C_{ai}) in the descending order.

IV. CONCLUSION

In this paper, two processes have been worked upon. Based upon user provided data, user area of interest is found. Following this strong friends are found using quasi-clique technique [1]. Finally, using both the data so obtained communities are being suggested. This extension of functionality gives more precise suggestion to the user regarding community suitable for him/her.

References

[1] *"Community Recommendation in Social Network Using Strong Friends and Quasi-Clique"* Approach Anjum Ibna Matin1 Sawgath Jahan1 Mohammad Rezwanul Huq2.

[2] Kadge, Sanam and Bhatia, Gresha, "*Graph Based Forecasting For Social Networking Site*", Communication, Information & Computing Technology (ICCICT) International Conference , 2012, , pp. 1-6.

[3] Monisha Kanakaraj1 and Ram Mohana Reddy Guddeti "*Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques*" Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015).

[4] Kamal Sutaria, Dipesh Joshi, Dr.C.K.Bhensdadiya Kruti Khalpada *"An Adaptive Approximation Algorithm for Community Detection in Social Network"*. 2015 IEEE International Conference on Computational Intelligence & Communication Technology.