International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-7, Issue-4, April 2017

A Friend Recommendation Algorithm Based on Social Network Trust

Lasheng Yu, Yu Yang, Xu Wu

***Abstract*— In recent years, the Internet into the Web2.0 era has brought the rapid development of social networks. But the huge user groups make it becoming increasingly difficult for the user to find like-minded potential friends. Therefore, almost all of the social networks provide recommended friend function based on the FOF algorithm, that is to say, friends of a friend will be recommended to the user. This approach mainly focuses on the relationship between users, but ignores the effect on which the user attributes will have on the formation for a new relationship between friends. Based on this, this paper proposes a new friend recommendation algorithm LWSNT (Local Walk based Social Network Trust), a limited steps walking algorithm based on the social network trust. The algorithm quantifies the user's attributes and serves as a reference index for the recommendation of friends, and introduces the trust mechanism based on the social network while taking into account the users' existing relationships, which makes the user more receptive to the recommendation result. Experiments show that the algorithm proposed in this paper is better than some similar algorithms if the target user's good friend information is enough.**

***Index Terms*— Friend Recommendation; Social Network; User's Attributes; Trust.**

I. INTRODUCTION

Social networks are generally based on the six-degree segmentation theory [1], a user expands his communication circle by making new friends. In recent years, the Internet into the Web 2.0 era has brought the rapid development of social networking services, for example, Facebook, Sina microblog ,QQ and other social networking platforms have millions or even hundreds of millions of large number of user groups, it becomes more and more difficult for users to find like-minded potential friends. Under this context, information recommendation based on social networking not only attracted great attention of researchers, but also have been widely used in social and online knowledge sharing sites. Compared with the traditional recommendation system, some of the recommendations based on social networks introduced a trust mechanism. Previous surveys from reference [2] have suggested that users prefer recommendations from friends rather than from the system. The studies in reference [3], [4] have confirmed the positive correlation between user interest similarity and user trust. you can increase the social networking site traffic and improve user dependency by providing users with satisfactory personalized service[5]. Therefore, to build a user's social trust network, and then to implement personalized friend recommended based on it will create great research value.

Lasheng Yu, School of Information Science and Engineering, Central South University, Changsha, China
Yu Yang, School of Information Science and Engineering, Central South University, Changsha, China
Xu Wu, School of Information Science and Engineering, Central South University, Changsha, China

II. RELATED WORKS

Jilin Chen, a researcher at GroupLens, explored the relationship between social interest and personal interest in information flow recommendations [6]. Le Yu proposed an adaptive social similarity calculation method based on matrix decomposition, which has been verified by Epinions [7]. Yu Haiqun and his collegue proposed a method based on user preference for social network recommendation [8]. Yuan T, Cheng J, Zhang X, et al. proposed a unified framework to properly incorporate the influence of social relationships into recommendation[9].

Chang WL, Diaz AN, Hung PCK mentioned that the trust value estimated will serve as a metric for filtering and sorting content of any kind based on the trustworthiness of the creator[10].Golbeck defines trust as: If user A determines that the behavior of user B will bring good results, then A will trust B. The rating of trust is generally measured by the degree of trust [11]. The similarity metric can be used as the calculation method of the trust between users in the social trust network. The method of local trust measurement in social network can be divided into the method based on node similarity and the calculation method of trust based on path.

The method based on node similarity takes the common neighbors or the node degree at both ends as the consideration. The similarity index of the common neighbors is CN(the number of common neighbors) [11], as shown in Equation (1), and the Jaccard metric [12], as shown in Equation (2). The premise of these two indicators is that if the number of common neighbors between two nodes is larger, the similarity between them will be greater.

Common-Neighbors:

$$sim(u,v) = |\rho(u) \cap \rho(v)| \tag{1}$$

Jaccard index:

$$sim(u,v) = \frac{|\rho(u) \cap \rho(v)|}{|\rho(u) \cup \rho(v)|} \tag{2}$$

The trustworthiness calculation method based on the path is mainly directed to the indirect trust users. The TidalTrust algorithm [15] is a kind of calculation method based on breadth first search proposed by Golbeck. Bouraga S, Jureta I, Faulkner S utilized a trust inference based algorithm to measure reputation score of each individual service, and subsequently trustworthiness of their composition[13].The idea of LRW (Local Random Walk) [14] can be expressed as follows: Given a graph and a starting point, the walker starts at random and moves to any adjacent node of the starting point. Adjacent nodes serve as new starting points. And the process will repeat and get a node sequence which is a random walk process. However, for an increasingly large social network, the global random walk has a large amount of computation which is almost computationally infeasible. Therefore, according to the six-dimensional theory [1], the LRW Friend algorithm only travels limited number of steps, that is, only consider the limited length of the path, generaly 2 to 6. The algorithm terminates in a relatively stable state without pursuing a steady-state distribution.

The Common-Neighbors and Jaccard metrics mentioned above are more concerned with the number of friends in common, and LRW Friend algorithm uses only the relationships among the points in the social graph, none of which used the attributes of the user to optimize the recommendation results. In this paper, we propose a new friend recommendation algorithm, LWSNT (Local Walking Based Social Network Trust), which combines the social graphs with user attributes to calculate the trustworthiness of social network users. The advantages of the algorithm are as follows:

(1) to quantify the user's attributes and use it as a reference index for friends' recommendation;

(2) to introduce the trust mechanism based on the social network diagram, and make the users more likely to accept the recommendation results;

(3) to get a better recommendation than the relevant algorithm in the case of sufficient user information, especially in the TOP-N recommendation can be achieved quite good results.

III. The Proposed Approach

A. Problem modeling

The relationship between users in a social network can be represented as a social graph of vertex sets and vertex pairs, where the vertices represent the users and the edges represent the social relationships between the users. In the social networks like Facebook, Renren and QQ, once A and B are connected, A is the friend of B and B is the friend of A, that is, the relationship between users is equal. The edge of the social graph is undirected, and it composing a undirected graph called the relational map[15].

Definition 1: According to graph theory, the social graph can be defined as $G = (V, E)$, where vertex set V represents all users in the social network, and non-directed edge set E represents the relationship pair existed in the social network. When there is a good friend relationship between the v_i and v_j, v_j is considered the adjacency point of v_i .The social graph G can be represented by the adjacency matrix $A = (a_{ij}) \in E$ .If there is a relationship between users, then a_{ij} = 1, otherwise a_{ij} = 0. This paper uses the form of adjacency list to store the social graph G.

In this paper, A user other than a friend of the target user in the social graph is referred to as a recommended user. The social map is shown in Fig.1:

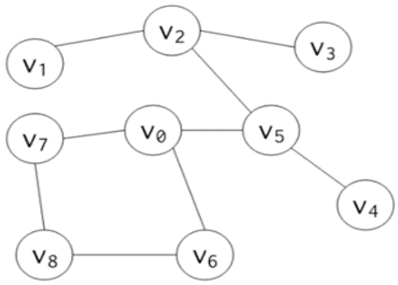

Fig.1 a simple social graph

Suppose in the social network, the user has c attributes [16], notated with $(x_1, x_2, \cdots, x_i, \cdots, x_c, \; i \in 1 \cdots c)$, such as sex, age, home, place of residence, hobbies, occupation, label, age of social accounts (from the date of registration) and their mutual friends, etc. Where interests can be subdivided, and each attribute has multiple values. These attributes may be independent or may have some relationship between each other. For example, a young male has a strong interest in sports, which identifies that age and gender will have a greater impact on a person's interest. If the interactions between attributes are considered, the problem will become more complicated, and it will get more and more difficult to calculate the degree of trust. Therefore, this paper will calculate the trust between the social network nodes based on independent users' attributes.

B. Trust Calculation

For convenience of description, the following definitions are given.

Definition 2: The target user's existing friend set is $F(v)$, size is $|F(v)|$.

Definition 3: The non-buddy nodes of the target user in the social graph form the recommended set named as the candidate set $C(v)$, and the size is $|C(v)|$.

Definition 4: The recommended list generated by the training set is $R(v)$, size is $|R(v)|$, which is the recommended result of the algorithm.

Definition 5: The user's real behavior list is $T(v)$, size is $|T(v)|$.

IV. User Attributes Quantification

The degree of influence of the i-th attribute on the trust degree of the target user is called "single attribute trust degree", represented by y_i.

There may be multiple values for a user attribute x_i. Taking gender as an example, assume that sex is the first attribute. Target user A has 55 male friends, 45 female friends, total friends $|F(v)|$=45+55=100. For a recommended male user B in the social graph, $y_1(x_1 = \text{male}) = 55/100 = 0.55$. The calculation of a single-attribute trustworthiness is shown in Equation (3):

$$y_i(x_i = values) = s_i(x_i = values)/|F(v)| \qquad (3)$$

Where $s_i(x_i = values)$ is the number of occurrences of the i-th attribute $x_i = values$ for all the friends of the target user. As can be seen from the formula 3, y_i is only related to the attributes of user's friends.

Let Y_{id} be the trust degree of the target user to the id-th recommended user. Equation (4) is the calculation formula of Y_{id} .

$$Y_{id} = \sum_1^c (f \times y_i(x_i = values)) \quad (4)$$

In the above formula, f is introduced to represent the average influence of each attribute, the initial value of $f = 1/c$. Considering that a certain attribute will have a greater impact on the recommended results, a constant z_i is used to enlarge the f value of the i-th attribute,and the sum of the influence of all attributes should always be 1,as shown in Equation(5).

$$\sum_1^c (z_i \times f) = 1 \quad (5)$$

So the new f value is calculated as shown in Equation (6).

$$f = 1/\sum_1^c z_i \quad (6)$$

Combining the above four formulas, the formula of trust degree will be defined as.

$$Y_{id} = \sum_1^c (z_i \times \frac{1}{\sum_1^c z_i} \times s_i(x_i = values)/|F(v)| \quad (7)$$

V. THE EFFECT OF CLOSELY RELATION ON TRUST DEGREE

In the social map mentioned in this paper, the target user can reach the adjacent node who is his friends, the node needed 2 steps to reach is a friend of friends, and the node taken 3 steps to reach is a friend of a friend's friend. In real life, the trust between users decreases with the number of walking steps. According to the six-dimensional theory, the number of walking steps k ranges from 2 to 6. In this paper, we introduce a trust attenuation factor $\alpha = \frac{k-t}{k}$, where t is the number of walking steps. The attenuation factor is combined with Equation 7 to obtain the final calculation formula (8). In this paper, the number of walking steps k = 4.

$$Y_{id} = \alpha \times \sum_1^c (z_i \times \frac{1}{\sum_1^c z_i} \times s_i(x_i = values)/|F(v)| \quad (8)$$

A. Algorithm Description

Combining the social graph and the new trust calculation method introduced above, this paper proposed a new friend recommendation algorithm.

The pseudocode of the friend recommendation algorithm LWSNT is shown in Algorithm 1. In step 3, the breadth-first search method is used to traverse the social graph and each point is traversed only once. Note that in the calculation of the trust degree Y_{id}, the result is related to z_i ,y_i and the current step number t, and only one parameter t is related to the node traversal. The breadth-first search itself can be regarded as a level order traversal, The number of access steps t to the current node must be the smallest of all traversal methods, while the degree of trust obtained is the largest of the current node. The purpose of using the heap is to reduce the time complexity of step 4. In the calculation process, The number of friends of the target user is $|F(v)|$, the number of nodes is n, the number of edges is m, and the number of edges traversed by the algorithm is m_v. The time complexity of Step 2 is $O(c \times |F(v)|)$. The time complexity of steps 3 and 4 is $O(m_v \times N log_2 N)$. The time complexity of Step 5 is $O(N log_2 N)$. From the previous analysis, the time complexity of step 2 and step 5 is relatively small, so the overall time complexity of the algorithm is determined by the time complexity of steps 3 and 4, that's to say, $O(m_v \times N log_2 N)$.

Algorithm 1 LWSNT

Inputs:

target user v_a , recommendation set size N (recommendation N friends to the target user), user information data set (including attribute information and friend relationship information $F(v_a)$ and $C(v_a)$). And z_i (may not input, the default value is 1)

Output:

Candidate recommendation set $R(v_a)$ consisting of Top-N recommended users for user v_a.

Step 1:

Create a social graph based on existing relationships.

Step 2:

Use the user attribute information of $F(v_a)$ to calculate the single attribute trust degree y_i of user attribute x_i using Equation(3).

Step 3:

Walking from v_a, using the breadth-first search method traverses the social graph. For the visited node $v \in C(v_a)$, the tag v has been accessed. the number of walking steps t_{id} is recorded and Equation(8) is used to calculate the trust degree Y_{id}. Each node only traverses once, that is, Y_{id} is calculated only once.

Step 4:

Create a small top heap with the size N and the heap with the key pair $< Y_{id}, id >$. According to the properities of the small top heap, if the new Y_{id} is larger than the heap key value of the trust Y_{id}, then pop-up the top elements of the heap and add a new key-value pair toensure that the Y_{id} of the elements in the heap is the current maximum N numbers.

Step 5:

The elements in the heap are the id and Y_{id} of the recommended set $R(v_a)$. The top of the heap elements will be deposited into a structure array in turn, and then output the elements in the array in reverse order.

VI. EXPERIMENTS AND ANALYSIS

A. Data Set and Preprocessing

To test the algorithm, this paper got data from Sina microblogging (www.weibo.com) and stored in the database. Sina microblogging, as the most active user microblogging platform, provides a complete set of open source interface, users can call the API directly to get data from the Sina servers. For the protection of user privacy, this paper only uses the public information of microblogging users. First of all, it selected 60 users as the target user, called V_1 , downloaded their user information,and then downloaded a total of 4896 users' information from the 60 users' friends, known as V_2. Finally, it downloaded the user information of V_2 friends', known as V_3, a total of 363,458 users. The data set is preprocessed, for every target user, random select 10% of their friends as the test set of the experiment, the remaining 90% as a friend set of the experiment.

B. Experimental measurement methods

A friend recommendation is a TOP-N prediction that focuses on whether a user will see the recommended content. This prediction generally provides users with a personalized recommendation list, using the prediction accuracy and recall rate to measure the recommended accuracy rate [20].

Recall is calculated as shown in Equation (9):

$$Recall = \frac{\sum_{v \in V} |R(v) \cap T(v)|}{\sum_{v \in V} |T(v)|} \quad (9)$$

The precision is calculated as shown in Equation (10):

$$Precision = \frac{\sum_{v \in V} |R(v) \cap T(v)|}{\sum_{v \in V} |R(v)|} \quad (10)$$

Recommended accuracy The F1 value is calculated as shown in Equation (11):

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

Obviously, the larger the F1 Measure, the better the performance of the algorithm.

C. Experimental results and analysis

The experiment selected the recommended set size N=1,2,5,10 and CN and Jaccard are compared with the LWSNT algorithm in this paper. Experiments were performed based on pre-processed data sets. There are 60 target users in the dataset, 14 of which have less than 50 friends and the remaining 46 have more than 50 friends. The experimental results are shown in Table. I and Table. II.

Table. I the F1 value of three kinds algorithms when the number of target users friends $|F(v_a)| < 50$

	N = 1	N = 2	N = 5	N = 10
LWSNT	0.3460	0.1623	0.1000	0.0629
CN	**0.4068**	**0.1677**	**0.1061**	**0.0643**
Jaccard	0.1086	0.0620	0.0482	0.0327

Table. II the F1 value of three kinds algorithms when the number of target users friends $|F(v_a)| \geq 50$

	N = 1	N = 2	N = 5	N = 10
LWSNT	**0.8123**	**0.5171**	**0.3846**	0.2353
CN	0.6429	0.5080	0.3820	**0.2474**
Jaccard	0.5286	0.4371	0.3229	0.2026

Table. I shows the performance of the three algorithms when the number of target users is small. From the target user friends to obtain the effective information is not much of the case, the performance of the three algorithms are not very good. In contrast, CN's performance is the best, LWSNT and its little difference, Jaccard's performance is relatively poor.

From Table. II and Table. I shows that the more friends of target users, the more effective information can obtained from friends, the better prediction accuracy of the algorithm. At this time, LWSNT has the highest accuracy, and only slightly worse than CN at N = 10. CN and Jaccard's performance has also been significantly improved.

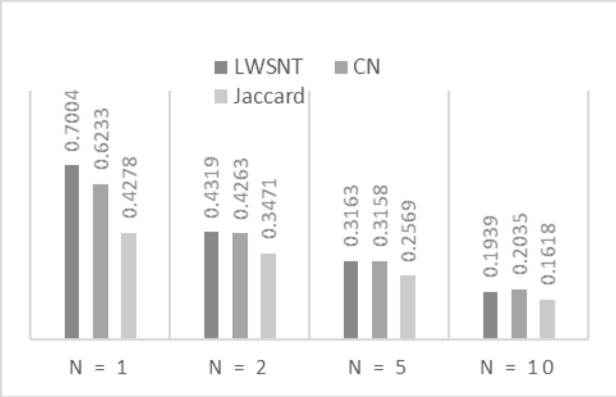

Fig.2 the F1 value of the three algorithms without regard to the number of target users

It can be seen from Fig.2 that LWSNT algorithm is slightly better than CN, Jarccard's performance is acceptable but is much worse than the other two algorithms.

Through the experimental study, the following conclusions can be drawn:

(1) CN, Jaccard and LWSNT of this paper are affected by the effective information. The more the friends of target users, the better the performance of the three algorithms.

(2) LWSNT algorithm in general have good performance but the recommended accuracy is higher at N = 1, 2, 5.

VII. Conclusion

This paper proposed a new friend recommendation algorithm based on social graph, LWSNT. Compared with other FOF algorithms, LWSNT utilizes the attributes of users to improve the accuracy of prediction. From the experimental results, the LWSNT algorithm presented in this paper is superior to the CN and the overall performance of Jaccard. This paper also finds that the performance of friend recommendation algorithm based on social graph is affected by the number of existing friends of the target user, and the more the number of existing friends is, the better the performance of algorithm is. In the case of a small number of friends, LWSNT algorithm performance is also acceptable. Of course, this study also received some restrictions. First of all, in the user attribute quantization, is assumed in a variety of properties completely independent of the case , the next study can consider the relationship between attributes. Second, the data set used in this experiment need to be extended. Finally, the time complexity of this algorithm is higher than that of CN and Jaccard, and the next research will focus on time efficiency optimization of the algorithm.

References

[1]. Milgram S. The small world problem[J]. Psychology Today, 1967, 32(2):185-195.
[2]. Sinha R, Swearingen K. Comparing Recommendations Made by Online Systems and Friends[J]. In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, 2001.
[3]. Ziegler C N, Golbeck J. Investigating interactions of trust and interest similarity[J]. Decision Support Systems, 2007, 43(2):460–475.
[4]. Bhuiyan T. A Survey on the Relationship between Trust and Interest Similarity in Online Social Networks[J]. Journal of Emerging Technologies in Web Intelligence, 2010, 2(4):291-299.
[5]. Deng S, Huang L, Xu G. Social network-based service recommendation with trust enhancement[J]. Expert Systems with Applications, 2014, 41(18):8075–8084.
[6]. Chen J, Nairn R, Chi E. Speak little and well: recommending conversations in online social streams[C]. Proceedings of the SIGCHI

Conference on Human Factors in Computing Systems. ACM, 2011:217-226.
[7]. Yu L, Pan R, Li Z. Adaptive social similarities for recommender systems[C]. Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011:257-260.
[8]. YU Hai-qun, LIU Wan-jun, QIU Yun-fei.Study on Second-level Network Recommendation Based on User's Preference [J]. Journal of Computer Applications and Software, 2012, 04 (04): 39-43.
[9]. Yuan T, Cheng J, Zhang X, et al. How friends affect user behaviors? An exploration of social relation analysis for recommendation[J]. Knowledge-Based Systems, 2015, 88(C):70-84.
[10]. Chang WL, Diaz AN, Hung PCK. Estimating trust value: A social network perspective[J]. Information Systems Frontiers, 2015, 17(6):1381-1400.
[11]. Golbeck J · Computing and applying trust in Web-based social networks〔D〕.University of Maryland，College Park，2005 ·
[12]. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et du Jura[J]. Bulletin De La Societe Vaudoise Des Sciences Naturelles, 1901, 37(142):547-579.
[13]. Bouraga S, Jureta I, Faulkner S. An empirical study of notifications'importance for online social network users[J]. Social Network Analysis and Mining, 2015, 5(1):1-34.
[14]. Liu W, Lü L. Link prediction based on local random walk[J]. Epl, 2010, 89(5):58007-58012(6).
[15]. Yigit M, Bilgin BE, Karahoca A. Extended topology based recommendation system for unidirectional social networks[J]. Expert Systems with Applications, 2015, 42(7):3653–3661.
[16]. Zhang Z, Liu Y, Ding W, et al. Proposing a new friend recommendation method, FRUTAI, to enhance social media providers' performance[J]. Decision Support Systems, 2015, 79:46–54.
[17]. Xiang Liang. Recommendation system practice [M]. People's Posts and Telecommunications Press, 2012,26-28.
[18]. Ma X, Lu H, Gan Z, et al. Improving Recommendation Accuracy with Clustering-Based Social Regularization[M]. Web Technologies and Applications. Springer International Publishing, 2014:177-188.
[19]. Ling Y, Guo D, Fei C, et al. User-based Clustering with Top-N Recommendation on Cold-Start Problem[C]. Proceedings of the 2013 Third International Conference on Intelligent System Design and Engineering Applications. IEEE Computer Society, 2013:1585-1589.
[20]. Rapečka A, Dzemyda G. A new Recommendation Model for the User Clustering-Based Recommendation System[J]. Information Technology & Control, 2015, 44(1).

Lasheng Yu, Vice professor in Central South University of China, ACM and CCF member, ACM/ICPC golden medal coach. He received the B.Sc. degree in Computer Science, the Master degree and a Ph.D. degree in Control Theory and Control Engineering from Central South University. He is the editor of Journal of Convergence Information Technology and Advances in Information Sciences and Service Sciences etc, he is also the reviewer for the journals such as Future Generation Computer Systems, journal of Parallel and Distributed Computing, Artificial Intelligence Review,etc. He has published at least 70 papers on Agent technologies or Algorithms, and has published 3 books. He has organized and implemented many projects which have created great achievements in the society.His main research interests include agent technologies and applications, structure and algorithm, smart computing etc.

Yu Yang, Master graduate of Central South University in China, majoring in computer science and technology.

Xu Wu, Master graduate of Central South University in China, majoring in computer science and technology