Find better friends?

- Re-configuration of personal networks by the neighborhood matchmaker method -

Masahiro HAMASAKI and **Hideaki TAKEDA**
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, JAPAN

Abstract

In this paper, we propose *Neighborhood Matchmaker Method* as a re-configration method for personal human network. Personal human network is important to realize the Semantic web because trust should be obtained and used with the personal network. Since a persnal network is usually ad-hoc network, it is necessary to extend and optimize it. Using the neighborhood matchmaker model, we can increase a new friend who is expected to share interests via all own neighborhoods on the personal human network. Iteration of matchmaking is used to optimize personal human networks. We simulate the neighborhood matchmaker model with the real data and the virtual data and compare the results by our model with those by the central server model. The neighborhood matchmaker model can reach almost the same results obtained by the sever model with each type of data.

Introduction

Anonymousness in web publishing is the double-edged sword in spreading web technologies across the world. We obtain the freedom to publish information without strict identification of ourselves. It is so attractive idea that web has accepted by people so quickly. But we are facing troubles because we have few methods to estimate creditability or reliability of web pages. Introducing concept of ”person” into web is the right way to make the web world more sound and natural. Now some proposals for it appear. For example, FOAF (Friend Of A Friend)[1] intends to provide a format for person and relations between persons. Then we can discuss the ”trust” of web as the matter of ”trust” of people (Jennifer, Parsia, & Hendler 2002). Personal networks or social networks are important to realize the trust layer of the semantic web because it is a source of trust and also a target of trust, i.e., trust can be obtained from personal network as well as trust can form personal network.

In this paper, we focus on how to re-configure personal networks to improve better relationship. Although it is said that any pair of people can be connected to the other by a short chain (Newman 2000), it is difficult to form a ”good” network. A personal network is usually ad-hoc network, i.e., most connections are established coincidentally. There is no guarantee that it is optimal. So we usually struggle to make our network better by adding or deleting friend connections. We call this problem ”personal network re-configuration problem”. Since the possibility of finding better friends is increased in the world of the Internet in comparison with the real world, ”personal network re-configuration problem” will be more serious. In this paper, we show an algorithm to this problem by using neighborhood people as matchmaker. It is a truly distributed method that does not need any central server, but it can archive reasonable results in fewer computation in comparison with the server case.

[1] http://rdfweb.org/foaf/

Related Work

There are some systems to capture and utilize personal human network in computer. Kautz et al. (Kautz, Selman, & Shah 1997) emphasized importance of human relations for WWW and showed done primary work for finding human relations, i.e., their system called ReferralWeb can find people by analyzing bibliography database. Sumi et al. (Sumi & Mase 2001a)(Sumi & Mase 2001b) supported people to meet persons who have same interests and share information using mobile computers and web applications. Kamei et al. supported to form communities using visualization relationship among participants(Kamei *et al.* 2001).

In these systems, they assume a group as a target either explicitly or implicitly. If the group is bigger, it is more likely to contain valuable persons to exchange information. However, we have to make more efforts with these systems in order to select such persons from a lot of candidates in the group. It is difficult for us to organize and manage such the large group. Therefore information exchanging systems should support methods that realize the above two requirements i.e. to meet and select new partners.

Furthermore, to support to meet some persons who have same interests is not enough to utilize personal human network. These systems supports to make personal human networks based interests people has. However, we need trusts among partners on personal human netwrok. We want to utilize personal human network based on interests sharing for valious utilization of information like information gathering. It is importants for valious utilization of information to share same interests and have a trust among partners. Generally, it is difficult for us to make a trust among new partners quickly on CMC. So, we need the method to make personal human networks based on relationship among persons. The proposed method is effective in this problem since this method extent personal human netwroks based on relationships which we had already.

Neighborhood Matchmaker Method

As we mention in the previous chapter, if we can need better relationship for information exchanging, we must meet and select partners more and more. It is a big burden for us, because we should meet all the candidates before we select them in advance. Since we do not know new friends before meeting them, we have no ways to select them. How can we solve this problem in our daily life? The practical way is introduction of new friends by the current friends. It is realistic and efficient because the person who knows both can judge whether this combination is suitable or not. Friends work as matchmaker for new friends. We formalize this ”friends as matchmaker” as an algorithm to extend and optimize networks.

The key feature of this approach is no need for central servers. The benefits of this approach are threefolds. The first is to keep spread of information minimally. Information on a person is transferred to only persons connected to her/him directly. It is desirable to keep personal information secure. The second is distributed computation. Computation to figure out better relationship is done by each mode, i.e., computers used by participants work for it. It is appropriate for a personal human network because we do not have to care the size of network. The third is gradual computation. The network will be converged gradually so that we can obtain the optimal network to some extent even if we stop the computation anytime.

Formalization

In this chapter, we introduce a model that can optimize networks by formalizing the method in our real life. We call that method *”Neighborhood Matchmaker Method (NMM)”* hereafter. Before explaining *NMM*, we define the network model for this problem. At first we define a person as a node, and a connection for information exchanging between people as a path. Here we assume that we can measure a degree of the connection between two nodes (hereinafter referred to as ”connection value”). Then, we can define that making a good environment for information exchanging is optimizing this network. In *NMM*, the network is optimized by matchmaking of neighbor nodes.

We need the following two conditions to apply *NMM*.

- All nodes can possibly connect to each other
- All nodes can calculate relationship between nodes connected to them

Under this condition, each node can change connections to others in order to connect better nodes autonomously. The behavior of a node is as follows.

1. Each node calculates connection values between its neighbor nodes.
2. If it finds pairs of nodes which have good enough connection values, it recommends them i.e., it tells each element of the recommendation pair that the pair is a good candidate for connection.
3. A node that receives recommendation decides whether it accepts the recommendation or not. If it accepts the recommendation, it adds the recommended node to its neighbor nodes.

We can optimize personal human network by iteration of these behaviors. Figure 1 shows these behaviors. In the next chapter, we test this method with simulations.

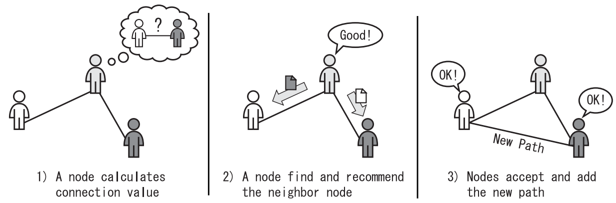

Figure 1: behavior of nodes

Experiments

The Procedure of the simulation

The main objective of the experiment is how our proposed method is useful to optimize a human network. We simulate optimization with *NMM* using the random data and the practical data.

Figure 2 shows the flow chart of this simulation. At first, we create nodes each of which has some data to represent a person. In this experiment, the data is a 10-dimensional vector and WWW bookmark taken by users. We initially put paths between nodes randomly. We fix the number of paths during simulation. It means that addition of a path requires deletion of a path. In this simulation the worst path in connection values is to be deleted when adding a new path. If all nodes cannot get a new path using matchmakers, the network is converged. Thus the simulation ends.

The Measurement of Evaluation

If it is possible to calculate connection values between all nodes at a place, we can create the best network definitely. We can measure how our method works by comparing the best network and the network generated by our method. We compare two networks in the following two ways. One is cover rate that is how much paths in the best network is found in the generated network. It means how much similar in structure two networks are. The other is reach rate that is comparison of the average of connection values between the best and generated networks. It indicates how much similar in effectiveness two networks are. These parameters are defined as the following formulas:

Simulation

Use the randam data

There are two parameters to control experiments. One is the number of nodes and the other is the number of paths. In this experiment, we set the size of nodes from 10 to 100 and the size of paths from the 1 to 5 times the number of nodes. The simulation is performed 10 times for each set of parameters, and we use the average as the results.

The graphs in Figure 3 and Figure 4 plot the average of cover-rate against turn. Figure 3 shows the results when the size of paths is fixed as thrice and Figure 4 shows the results when the size of nodes is fixed to 60.

The algorithm does not grantee that the network will be converged. However, since all graphs became horizontal, all networks were converged using matchmaking. We furthermore can find that the average of cover-rate and the turn of convergence are dependent on the size of nodes and paths.

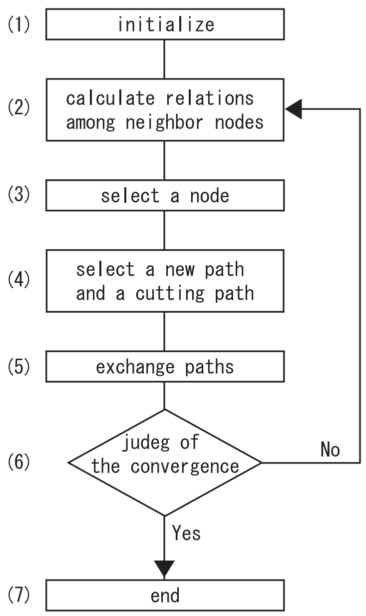

Figure 2: Flow chart of simulation

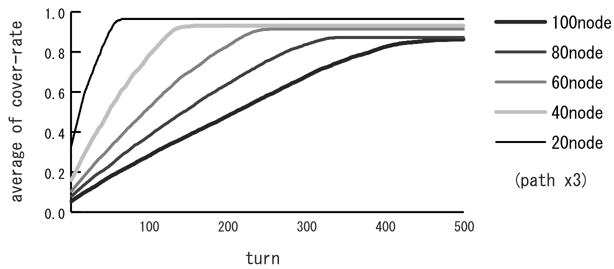

Figure 3: Cover-Rate and nodes in the random data

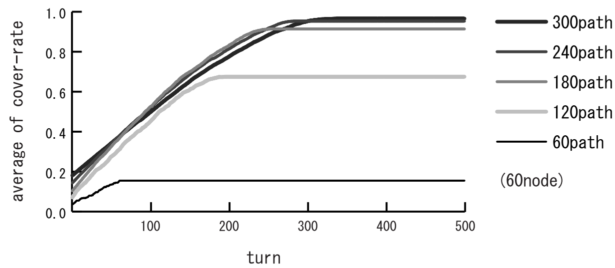

Figure 4: Cover-Rate and paths in the random data

$$cover\ rate = \frac{|\{P_{current} \cap P_{best}\}|}{N}$$

$$reach\ rate = \frac{\sum_{l=1}^{N} f(p_l | p_l \in \{P_{current}\})}{\sum_{m=1}^{N} f(p_m | p_m \in \{P_{best}\})}$$

p	:	a path
N	:	the size of paths
$\{P\}$	:	a set of paths
$\{P_{best}\}$	:	the best set of paths
$\{P_{current}\}$	:	the current set of paths
$f(p)$	:	a value of path

We observed similar results on reach-rate. The difference is reach-rate is less dependent on size of paths and nodes.

We also examine the relevance between the size of network and the turn of convergence. After iteration of simulation varying size of nodes and paths, we obtain the graph in Figure 5 plots the average of convergence turns against the size of nodes. This graph indicated that the turn of convergence increases linearly when the size of nodes increases. In this simulation, only a single one node can exchange paths in a turn, so the times of exchanging per node do not became so large.

The number of turns of convergence is linearly increased against the size of nodes. In this simulation, only one node can exchange a path to own neighborhood nodes in a turn.

Let me estimate order of the algorithm roughly. When the average of the number of neighborhood nodes is r, this algorithm calculates connection values $2r$ times in every turn. When the size of nodes is N and the number of turns of

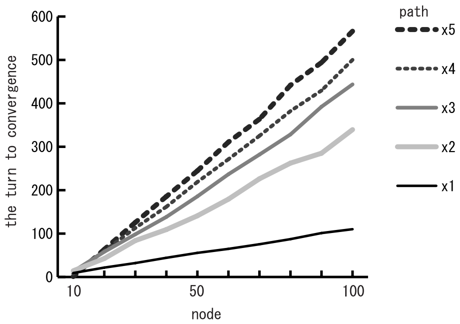

Figure 5: Average of Convergence Turn

convergence is kN according to Figure 5, the calculation times to converge is $2rkN$ using *NMM*. In the centralized model the calculation times is N^2 because we have to calculate connection values among all nodes. Since r and k are fix value, the order is $O(N)$ using *NMM*. It is less than $O(N^2)$ using the centralized model.

Use the practical data

We also used the practical data generated by people. We use WWW bookmarks to measure connection values among people. Users always add a web page in which she/he is interested, and organize topics as folder in WWW bookmark. So it can be said that WWW bookmark represents the user profile. In this simulation, we need to calculate relationship between nodes. We use a parameter called *"category resemblance"* such as a value of relationship between nodes (Hamasaki & Takeda 2001). This parameter is based on resemblance of folder structure of WWW Bookmark.

We asked twelve persons to submit their WWW bookmark files to the system that can calculate relationship between WWW bookmarks. We set up the size of nodes as 12 and the size of paths as 12, 16, 20, and 24. The simulation is performed 10 times for each set of parameters, and the average as the results. We examine the average of measurements and convergence turns. We found that there is the similar tendency with the random data. These results indicate that the network could be optimized in the practical data.

Conclusion

In this paper, we propose the way to obtain a new person who is a partner for exchanging information and proposed a method called *"Neighborhood Matchmaker Method (NMM)"*. Our method use collaborative and autonomous matchmaking and do not need any central servers. Nevertheless, by examining our experiment results, the optimal personal human network can be obtained. In this simulation we need the number of paths that is 2 to 3 times of the number of nodes and the number of turns that is 1.5 to 2 times the number of nodes in order to optimize the network sufficiently.

This method does not need the centered server. This feature has three advantages for Personal Human Network. The first is to make personal information free from the risk of central servers in the viewpoint of security. The second that we can deal with any size of groups, because it calculates relationship among people without collecting all data at the server. It is possible to assist bigger groups that are more likely to contain valuable persons to exchange information. The third is that we can use this method easily and quickly, so that we can assist to form dynamic and emergent communities that are typical in the Internet.

A further direction of this study will be to develop a system using this proposed the method and investigate effectiveness for it in real world.

References

Hamasaki, M., and Takeda, H. 2001. Experimental results for a method to discover human relationship based on www bookmarks. In *Proceedings of the Knowledge-Based Intelligent Information & Engineering Systems (KES2001)*.

Jennifer, G.; Parsia, B.; and Hendler, J. 2002. Trust networks on the semantic web. Submitted to WWW2003.

Kamei, K.; Jettmar, E.; Fujita, K.; Yoshida, S.; and Kuwabara, K. 2001. Community organizer: Suppoting the formation of network communities throuh spatial representation. In *Proceedings of the 2001 Symposium on Applications and the Internet (SAINT2001)*.

Kautz, H.; Selman, B.; and Shah, M. 1997. Referral web: Combining social networks and collaborative filtering. *Communications of the ACM* 40(3):63–65.

Newman, M. 2000. Models of the small world - review. 819–841.

Sumi, Y., and Mase, K. 2001a. Collecting, visualizing, and exchanging personal interests and experiences in communities. In *the 2001 International Conference on Web Intelligence (WI-01)*.

Sumi, Y., and Mase, K. 2001b. Digital assistant for supporting conference participants. In *An Attempt to Combine Mobile, Ubiquitous and Web Computing(Ubicomp 2001)*, 156–175.