Trust Network-based Filtering to Retrieve Trustworthy Word-of-Mouth Information

Hiromitsu Kato, Yoshinori Sato, Takashi Fukumoto, Koichi Homma, Toshiro Sasaki, and Motohisa Funabashi

Systems Development Laboratory, Hitachi, Ltd.
1099 Ohzenji, Asao-ku, Kawasaki 215-0013, Japan
{hkato, y-satou, fukumoto, homma, t-sasaki, funa}@sdl.hitachi.co.jp
http://www.sdl.hitachi.co.jp/

Abstract. We propose a method to retrieve trustworthy information from a word-of-mouth community space that could include inappropriate information. In a ubiquitous information society, high anonymity could cause some problems concerning human rights infringements, such as slander and invasion of privacy. On the other hand, unless anonymity is guaranteed at a certain level, free expression and speech would not be possible and may lead to an inactive information society. As a countermeasure to such problems, information rating-based filtering was examined in previous work. However, some difficulties such as the infringements of freedom of expression and the objective authenticity of the rating results have been pointed out. In this research, we pay attention to the reliance on relationships between people, and construct a model in which information is valued when it comes from people who can be relied on. Also, we aim to construct a framework for evaluating the trustworthiness of information by forming a community that is generally known as the "web of trust" model.

1 Introduction

In this Internet era, a wide variety of information in the world has become more accessible to many people. Once barrier-free information is achieved after the development and spread of ubiquitous information technologies, it is expected that the next generation society, with anyone able to interact with anything freely at anytime, will arrive.

In the future IT society, the real world and information will become deeply connected, and thus it will be possible that images will be able to have attributes that can correspond to the actual position and time of where and when the picture was taken with a camera. As the real and information worlds become linked, it will become possible to execute services according to the user position and use the circumstances of the real world as the context.

However, it is probably inevitable that the free submission and retrieval of information will be socially restricted. For example, the ease of information handling and the high anonymity in the current Internet makes it easy to send

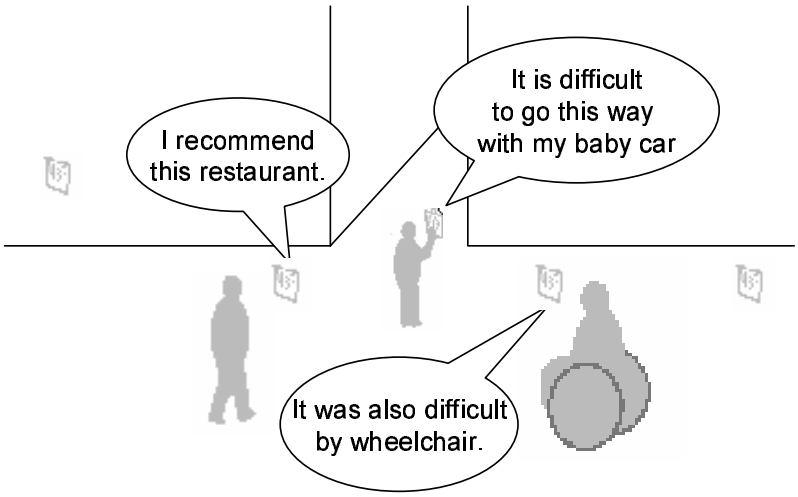

Fig. 1. This figure shows an image of a Ubiquitous Bulletin Board. The system allows people to put a virtual memo anywhere they want. It is possible to put it with a photo taken by a camera attached to a mobile phone. Since the memo is also linked to the location information, a collection of UBB information can be drawn on a map.

and view material that infringes human rights, such as slanderous and private information. While free expression is likely to reduce if there is no guarantee of anonymity, which would lead to an inactive information society, anonymity could cause behavior that infringes human rights.

In a ubiquitous information society, we need to consider more deeply how to prevent invasions of privacy. The development of a secure system platform to suppress the circulation of disinformation and slanderous information is expected, since this will make it much easier to collect and trace user's behavior history and preferences.

The goal of this research is to provide a mechanism for accessing information supported by a community of people that is trusted by the user to evaluate the disinformation and slanderous information that is circulating in the ubiquitous information society. We also consider a means to protect privacy when private information is needed to evaluate the trustworthiness of the information.

2 An Application Example - Ubiquitous Bulletin Board -

Before discussing system design, let us assume a typical service in the ubiquitous information society. In this paper we assume a virtual memo to be word-of-mouth information, which we can stick to a specific place so that other people can see it. We call this application the "Ubiquitous Bulletin Board" (UBB) as shown in Fig. 1.

However, this system may have some serious consequences and result in privacy invasion, such as

- circulation of false information,
- infringements of portrait rights and privacy of a specific individual,
- circulation of slanderous information about a specific individual or organization, and
- circulation of information which incites public disorder or insults customs.

Although these issues do not only originate from the technical issues but also from human behavioral and ethical issues, technology should help users to keep away from these kinds of unwanted information. In this research, we aim to provide a means to prevent us from accessing unsuitable information, even though it is difficult to oppress this information in a free world. We also aim to offer a means of supporting smooth access to useful information at the same time.

3 Assumptions

Taking into account some ethical and psychological considerations, we assume that the following system requirements:

- All information lacks in objectivity and authenticity.
- Users do not expect thoroughly rated information. It is common for information to be unrated.
- Users rely on ratings by responsible people or authorities who you can be trusted. However, users cannot rely on whether the information rated by such people is trustworthy.
- The end user does not arrange any filtering policies explicitly and manually.

4 Previous works

The main technical requirements for achieving the above objectives can be summarized as the following two items:

(1) Filtering of inappropriate information
(2) Prevention of release of unnecessary private information

In this section, let us recap the previous works.

4.1 Information Filtering

Content-based Filtering The techniques for content filtering can be classified as the following three categories:

- Use of blacklists
- Use of whitelists

– Self-labeling and filtering

For techniques using blacklists, the third party agency checks the contents of the Web page, and picks out any inappropriate content as a blacklist, which restricts the content seen by users.

The technique using whitelists lists only the desirable content, the opposite of the blacklist, and only these contents can be accessed. Since the available services are quite limited, this approach may be effective only for educational techniques for young children.

The rating method that labels the contents of a Web page is performed by the content provider using a prepared checklist. This is generally called "self-rating", and Platform for Internet Content Selection[3] (PICS) is a standard for such labeling. PICS offers a mechanism to show the ratings of Web contents as judged by parents and teachers. Then, children and pupils can view the desired contents, while being successfully kept away from inappropriate content. The places where self-rating and the user filtering using PICS is effective is limited to organizations like schools since:

– there is no obligation for website operators to perform self-rating,
– there are no penalty regulations for lying about a self-rating, and
– even incompetent users must set a filtering policy.

Collaborative Filtering The technology which extracts useful items from a lot of information using user knowledge is called "social filtering"[5] or "collaborative filtering"[4]. This technology is often used in a recommendation system that suggests items with the highest appraisal in the user group that matches the target user. If the hypothesis that items which similar users have used is useful to the target user is true, it can be applied to authenticity appraisal for word-of-mouth information. The value of the rating can be obtained from the explicit ratings by the user or the implicit history of the user's behavior, such as choice of bookmarks and frequency of the service use.

However, when collaborative filtering is used, obtaining the past records of many users is necessary to construct a reliable user model. As a result, this tends to take a long time. In addition, since it is necessary to collect private information from the user to construct the user model, the system should be developed with a consideration of privacy protection.

Ranking Page et al[7] proposed a method to evaluate web pages using hyper-linked relations between pages. They used a hypothesis that the page which is linked by good pages is also a good page, and modeled the relationship of web pages as a directed graph based on the link information between web pages.

This concept can be extended to human-to-information relation modeling. Namely, the reliance network model that considers the reliability of the user who offered the word-of-mouth communication information.

The reliability appraisal that uses the information social network and includes user information aims at improving the quality of the appraisal by taking

into account the user. However, it is necessary to consider privacy protection when it is difficult to obtain the network structural information because private information is included in the links.

Use of human networks To prevent the web bulletin board from becoming an illegal zone, Umeki et al[8] proposed the system called VOTE. In this the participants evaluate the contributions and contributors by voting. The rights to control the bulletin board are assigned to the contributors according to the received positive votes. However, the framework of VOTE is limited to the Web bulletin board.

Takeuchi et al[9] proposed the Human Network based Filtering (HNF) system to measure the value of information based on the trustworthiness of the person who has sent or forwarded information. Since the HNF system is designed based on the word-of-mouth model of human society, only information the relevant users (neighbors) appraised is forwarded. This means that unnecessary information is filtered according to the judgment of the contiguity of users. However, since HNF is designed to be a broadcast-type information sharing system, it does not offer the function to get appropriate information at appropriate times. On the other hand, Ito et al[10] developed Word-of-mouth-Assisting Virtual Environment (WAVE) as a test to achieve the word-of-mouth model in the e-community. Since the users can obtain the required information from their neighbors in WAVE, the availability of the distributed information is limited. This limitation is one area that needs to be improved.

Chen et al[11] proposed the system called Poblano, which can be used to search and rank information based on the confidence evaluated by the relevant users in the peer-to-peer (P2P) distributed information sharing system. In Poblano, we suppose that each peer has confidence in the information ($CodatConf$, where“ codat ”means codes and data) and confidence in other peers ($PeerConf$).

4.2 Privacy Protection

Informed consent The other big topic is privacy protection. One of the major approaches here is ”notice and consent” or informed consent. Namely, the presentation of a privacy policy and the agreement of the user are necessary before submitting the personal information. This model is implemented in P3P (Platform for Privacy Preferences), which is used in Web transactions to prevent the outflow of private data.

As for P3P, the privacy policy presented the observance possible security basis and execution of management of use are necessary to observe security and management issues of the enterprise website providers. In addition, the user interface and the system construction [which make that to the end user it offers in understanding possible shape in regard to privacy policy and preferences, sets information possible are necessary[12].

Anonymous communication To protect privacy, several anonymous communication systems are proposed to give the individuals anonymity. Chaum's MIX-net[13] is prominent as an anonymous conversion for one-to-one communication networks, and is used in applications such as electronic polls.

Reiter et al[14] proposed Crowds as a technology to provide anonymity by using user groups. Crowds makes it difficult to identify the sender of information by selecting the requestor at random. Clarke et al[15] extended the idea of Crowds and developed Freenet, which is aimed at improving anonymity, confidentiality, and efficiency of data storage.

5 System Requirements

In this paper, we assumed that the polling action in VOTE is comparable to the evaluation feedback in Poblano. Hence, we propose trust network formation based on the Poblano model. However, to resolve the current problems related to Poblano, we argue the system requirements are as follows.

Trust Relationship Acquisition Let us assume a model which implies that those to whom we send messages more frequently are more reliable. In this case, after the frequency of the message transmission is analyzed from the real (physical interactions) and virtual (electronic interactions) communication logs, we can obtain a peer confidence table. It can also be understood which keyword group each peer belongs to by using a keyword matching technique.

Anonymity and Confidentiality If we consider that frequent communication with other people often involves highly confidential personal data, such private data should not be disclosed, even to trusted people. However, the original Poblano model opens the peer identity and confidence information to the public. For example, as shown in Fig. 2, let us assume that Peer-C, which holds the content Peer-A requested, is discovered by Peer-B, and that the content is transferred though the peer-to-peer communication between A and C. In this case, the accumulated value of the peer confidence from the starting peer to the previous peer (Peer-B) is disclosed to the people on the multi-hop path. Then, if the accumulated value is no bigger than the maximum value, Peer-A would suspect that the peer confidence of Peer-B for Peer-C could be open, especially after the similar searches result in the same confidence value for this B-to-C trust link.

We classified the P2P network systems into the following three categories.

(a) Publicly identifiable
(b) Locally identifiable
(c) Anonymous

The typical P2P network systems are categorized in Type (a). As we reviewed in Section 4.2, there are also several anonymous P2P systems, which are grouped

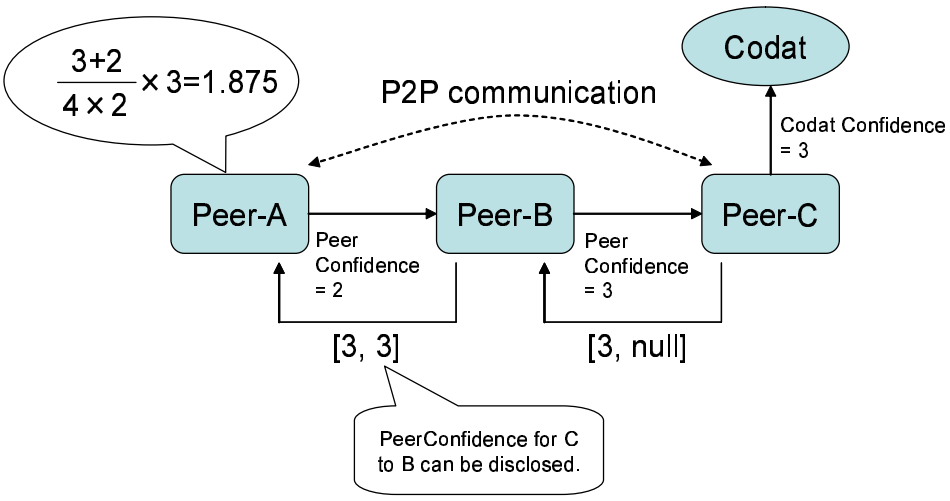

Fig. 2. This figure shows a diagram of trust propagation in the original Poblano. The notation [X, Y] means that X is the codat confidence of the end peer and Y is the sum of the peer confidence on the multi-hop trust path from the requester to the provider (end peer). The peer confidence for Peer-C to Peer-B could be disclosed to Peer-A on the way to the data transfer [X, Y].

into Type (c). However, trust value cannot be evaluated in the complete anonymous system. Hence, Type (b) is suitable for our trust network model to enable us to evaluate trust value and to conceal it from other users. Since the original Poblano model can be considered as Type (a), it is necessary to improve the Poblano model for Type (b).

Taking this point into account, we propose to re-define the formulation for the trust model as Eq.(1) as a counter measure to the trust disclosure problem.

$$CodatConf_{path}(T) = \prod_{i=1}^{n} \left(\frac{PeerConf(P_i)}{MaxValue} \right) \times CodatConf(T) \qquad (1)$$

where $PeerConf(P_k)$ is the trust value of the k-th peer, P_k, evaluated by the (k+1) peer, $MaxValue$ is a maximum value of $PeerConf(P_k)$, $CodatConf_{path}(T)$ is the trust value of a target, T, for a single trust path, and $CodatConf(T)$ is the trust value of T evaluated by the provider peer. Using this formula, k-th peer just calculates

$$CodatConf_k(T) = \frac{PeerConf(P_{k-1})}{MaxValue} \times CodatConf_{k-1}(T) \qquad (2)$$

for the $CodatConf_{k-1}(T)$, which is forwarded by the neighbor (k-1)-th peer P_{k-1}. At the end, the requestor peer can obtain $CodatConf_{path}$. Since it cannot be disclosed whether $CodatConf_{k-1}$ is evaluated by the P_k or the product of the forwarded values from the upstream of the trust path, as shown in Fig. 3, the Crowds-type can achieve anonymity.

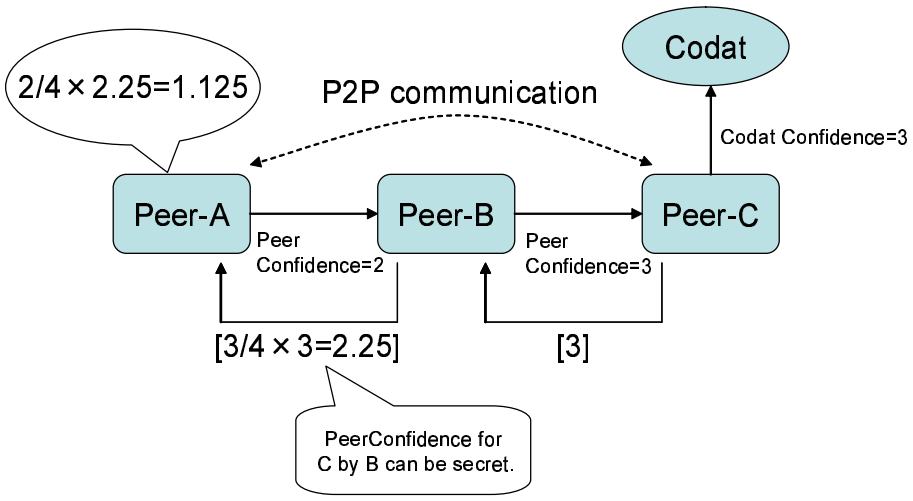

Fig. 3. This figure shows a diagram of the revised trust propagation model that we propose. Since the transferred information is the value calculated from Eq.(2), Peer-A receives a value of 2.25 from Peer-B. In this case, Peer-A cannot obtain the peer confidence of Peer-B for Peer-C.

Community Restriction When information sharing is limited within the restricted trust network, it is necessary to check the platform environment of the communication partner so that it cannot forward data to a third party, or make a contract on equal terms.

Push and Pull Distribution Although the information search is a Pull-type service since end users search on the trust network community, a pseudo-Push-type service will be useful for users for services like event-based coupon distribution.

Context-based Search The system should be able to search not only by keyword but also by some context, such as location information. For this objective, the information should be categorized based on context as well as a series of keywords.

6 Summary and Future Work

As a countermeasure against the flood of inappropriate information in a ubiquitous information society, we studied a trust network model to filter such information. In our research the trust network can be mapped to real human trust relations. In this paper we discussed related previous work and clarified their difficulties. Also, we took into account the problem of privacy protection, and listed the system requirements to resolve these problems.

For future work, we need to continue studying the following issues:

- We expect that a trust network will be a "small world" network[16]. Hence, the scalability for information retrieval needs to be studied since it will be scalable by the small-world effect.
- Even if we apply our trust model to information filtering, inadequate information will still be circulating among the "dark side" communities, where people want to collect that type of inappropriate information. Since technology cannot provide a complete solution, we also consider non-IT countermeasures, such as media literacy education.
- When people participate in the UBB, they will be eager to retrieve information. However, since there is little incentive to contribute information, it is predicted that the community will not be active. We may need to check the sociological analysis to overcome this difficulty.

7 Acknowledgement

This research is conducted as a program for the "Promotion of Leading Research" for the Special Coordination Funds for Promoting Science and Technology by the Ministry of Education, Culture, Sports, Science and Technology.

References

1. Mizutani, M.: Information Ethics: An Introduction. IPSJ SIGNotes Electronic Intellectual Property, Vol.99, No. 11 (1999) 127–134 (in Japanese)
2. Kiesler, S., Siegal, J., Mcguire, T. W.: Social Psychological Aspects of Computer-Mediated Communication. American Psychologist, Vol.39, No.10 (1984) 1123–1134
3. Platform for Internet Content Selection (PICS) http://www.w3.org/PICS/
4. Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl J.: Item-Based Collaborative Filter-ing Recommendation Algorithms. In Proc. of the 10th International world Wide Web Conference (WWW10) (2001) 285–295
5. Shardanand, U., and Maes, P.: Social Information Filtering: Algorithms for Automating 'Word of Mouth'. In Proc. of CHI '95 (1995) 210–217
6. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Reidl, J.: GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference, New York, NY (1994) 175–186
7. Page, L., Brin, S., Motwani, R., and Winograd, T.: The PageRank citation ranking: Bringing order to the web. (1998)
8. Umeki, H., Shimogoori, N., Yokota T.: Supporting Network Community Formation. ISPJ SIG Notes Information Media, No.037-005 (2000) 25–30 (in Japanese)
9. Takeuchi, S., Kamahara, J., Shimojo, S., Miyahara, H.: Human-Network-based Filtering: The Information Propagation Model based on Word-of-Mouth Communication. Proceedings of the 2003 International Symposium on Applications and the Internet (SAINT-2003) (2003) 40–47
10. Ito, Y., Yoshida, M., Numao, M.: New features of the Word-of-mouth Assisting Virtual Environment. IPSJ SIGNotes Intelligence and Complex Systems, No. 128-028 (2002)

11. Chen, R. and Yeager, W.: Poblano: A Distributed Trust Model for Peer-to-Peer Networks. Sun Microsystems Technical Paper (2000) http://www.sun.com/software/jxta/poblano.pdf
12. Langheinrich, M.: A Privacy Awareness System for Ubiquitous Computing Environ-ments. In: G. Borriello, L.E. Holmquist (Eds.): 4th International Conference on Ubiquitous Computing (UbiComp2002), Springer-Verlag LNCS 2498 (2002) 237–245
13. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 24(2), Feb. (1981) 84–88
14. Reiter, M. K. and Rubin, A. D.: Crowds: Anonymity for web transactions. ACM Transac-tions on Information and System Security 1(1), November (1998) 66-92.
15. Clarke, I., Sandberg, O., Wiley, B., and Hong, T. W.: Freenet: A distributed anonymous information storage and retrieval system. In Workshop on Design Issues in Anonymity and Unobservability, ICSI, Berkeley, CA, USA. (2000) 311–320,
16. Watts, D. J. , Strogatz, S. H.: Collective dynamics of ‘ small-world ’ networks. Nature, Vol. 393, No. 4 (1998) 440-442