Evaluation of Characteristics of Word-of-Mouth Communication Forms for Predicting Information Propagation on the Internet

Susumu Takeuchi[†] **Masanori Akiyoshi**[†] **Norihisa Komoda**[†]
[†]Graduate School of Information Science and Technology, Osaka University
{stakeuti,akiyoshi,komoda}@ist.osaka-u.ac.jp

Abstract

Since information is easily and widely spread on the Internet recently, detecting undesired information that affects many users prior to its propagation is necessary. By focusing on word-of-mouth communication forms of information, which are constructed by the users' subjective impressions and actions, possibilities of predicting information propagation based on its forms are verified by performing an experiment with real users. As a result, the average number of users whom information passed through, i.e., propagation distance, of interesting or important information tends to larger. This result suggests that evaluating characteristics of information by its propagation forms may be possible.

Keywords: Information Propagation, Word-of-Mouth Communication, Average Propagation Distance

1 Introduction

Currently, users can obtain information by utilizing the devices that allow the users to connect to the Internet, e.g., personal computers, and cellular phones. Various kinds of services are also available on the Internet that enables the users to send information without difficulties. For example, blog (web log) encourages the users to publish their opinions about specific news or dairies without knowledge of HTML tags. Consequently, various kinds of information are sent by varieties of users, and such information tends to be propagated throughout the Internet.

Information, which is propagated to many users, sometimes includes criticisms, slanders, discriminatory comments, or leaked personal information. Propagation of these kinds of information affects extremely the person or the company concerned. Meanwhile, even if information is harmless by itself, propagation sometimes causes a problem such as chain e-mail. Therefore, detecting information propagation without considering the content of it prior to its propagation is significant for maintaining healthy and reliable communication on the Internet.

In this paper, by focusing on word-of-mouth communication forms of information, possibilities that estimating propagation by the traces of its circulation are estimated. For investigating the possibilities, an experiment based on word-of-mouth communication is performed by real users. When propagation can be estimated by its forms, undesired propagation may be detected independent of its content and is expected to be prevented.

The rest of the paper is organized as follows. Section 2 explains the features of information propagation on the Internet. In Section 3, a criterion is proposed to evaluate propagation forms. Section 4 describes the details of experimental environment and results of the evaluations. Finally, Section 5 concludes the paper.

2 Information propagation on the Internet

2.1 Circulation features in existing communication systems

To address a detection or prediction of information propagation, features of circulation of information should be identified. Accordingly, communication systems which are used on the Internet is classified as shown in Table 1.

In case of general web site or community services such as BBS (Bulletin Board System), SNS (Social Networking Service), blog site, metaverse the general public share the same "place" to communicate so that information sometimes propagates among many users. Estimating the number of users who receive such information is difficult because the number of receivers depends on the contents of information.

Meanwhile, in case of e-mail or instant messaging, information is transmitted to specified target users so that information does not prop-

Table 1. Classification of communication systems

	Target users are:	
Number of target users	Specified	Unspecified
few	E-mail, instant messaging	
many	Mailing list, groupware	BBS, blog, metaverse

agate to outsiders. In case of a mailing list or a groupware, the users communicate with multiple users, but a community is closed so that information may not propagate to other users who do not join the community.

2.2 Features of problems of information propagation

In the systems which support communication between specific users, information does not propagate directly to other users as described in Section 2.1. However, information is often forwarded by a user who is one of the limited users. Thus, information sometimes propagates among many users by word-of-mouth communication like chain e-mail. Meanwhile, even if information can be obtained by many users such as on the Web, which is a form of propagation, the propagation of information does not always cause a problem. Because the problem is occurred when a receiver is willing to take action on it, e.g., transmits it to other users with the receiver's comment, according to its importance or interest. Obviously, this action is a basis of word-of-mouth communication.

Therefore, information propagation should focus not on the number of receivers but on how the receivers feel and transmit the information. In other words, information can be considered to *propagate* when circulations in word-of-mouth communication are frequent.

2.3 Related Work

In order to identify information propagation, two approaches can be considered. One is "trend finding" approach which observes hot topics by counting the number of same word or similar contents ([1] etc). The analyzers may be possible to identify undesired information from these topics, but it is too late to prevent its propagation because hot topics have already propagated.

The other is "web mining" approach which sets keywords, e.g., name of a product, company, or place, relates to information that specific people do not desire to be propagated among many users, and collects the pages which contain such keywords ([2] etc). Consequently, the analyzers may be able to find undesired information that includes the keywords. In contrast to the trend finding approach, a gathering system may alert the analyzers when information, which includes the keyword, is sent by someone prior to its propagation. However, when an abbreviation or euphemism is used instead of the specified keyword itself, detecting by the keyword may become difficult and force the analyzers to set multiple synonyms as keywords. Moreover, a provider, which serves a community service such as BBS, SNS, blog site, or metaverse, cannot assume a keyword which indicates undesired information in advance, because undesired keywords emerge day by day.

Besides, if specific information can be detected by the keyword, deleting such information has a risk to cause *flaming*. Thus, a deletion of undesired information should be based on the estimation of information propagation, i.e., undesired information should be deleted when the information is estimated to propagate among many users.

3 Predicting word-of-mouth propagation by its forms

3.1 Propagation of word-of-mouth communication

As described in Section 2, word-of-mouth communication is significant for predicting information propagation. The followings are considered natures of word-of-mouth communication on the Internet.

- Information is transmitted between the real users.
- Receivers of information are indicated by a sender or transmitter of the information.
- Contents of information are not easily altered fundamentally, but comments are

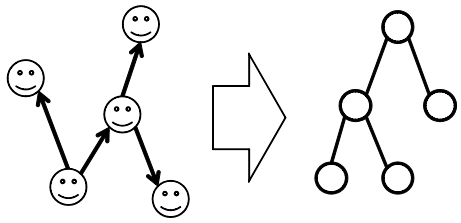

Figure 1. Example of a propagation tree of information in word-of-mouth communications

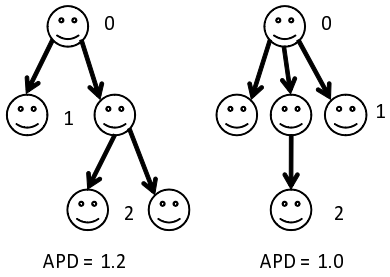

Figure 2. Correlation between Average Propagation Distance and specifications of propagation forms

sometimes appended.

In word-of-mouth communication, information is transmitted and propagated only when a transmitter judges the information is worth to be shared with the receivers. From the social sciences point of view, information propagates by the subjective impressions of people in the real world, such as importance and vagueness of information [3; 4], or uncertain of information [5]. Thus, if subjective impressions of the users can be obtained, predicting information propagation may be possible. However, the number of users and information is extremely huge, prediction based on subjective impressions of the users is impossible.

3.2 Word-of-mouth communication forms

In word-of-mouth communication, a propagation tree is constructed based on the relation of senders and receivers. Figure 1 shows an example of propagation tree. Each edge indicates the relationship between a sender and a receiver, and each node indicates a sender or a receiver. User forwards information when the user thinks it is worth to be shared, so the tree can be considered to include subjective impressions of the users. Thus, difference of the forms is expected to express the difference of nature of information.

The parameters and the features of propagation forms, which can be considered to obtain from a propagation tree, are the followings.

- Propagation distance of information

  Information, which is forwarded by multiple users, can be considered significant among the users. Thus, propagation distance is considered to relate to subjective evaluations of the users.

- Number of propagated users

  The number of propagated users from a specific user may indicate how information is significant for the user. Thus, the number of propagated users is considered to relate to subjective evaluations of the user as well.

3.3 Average propagation distance

For evaluating the features of information based on the parameters described in Section 3.2, the measurement of Average Propagation Distance (APD) is proposed in [6]. The value of APD indicates the average number of users whom information passed through. The APD of specific information i is calculated by

$$apd(i) = \frac{\sum_{\forall u \in R} l(i,u)}{|R|} \quad (1)$$

where R indicates a set of users who receive i, $l(i,u)$ indicates a propagation distance (0 in case of a sender) of i when a receiver $u \in R$ receives, and $|R|$ indicates the number of receivers.

As shown in Figure 2, even if the number of receivers of specific information is the same, the value of APD differs based on its propagation forms. When the value of APD is large, information is significant among many users because the number of users who forward it is large. On the contrary, when the value of APD is small, information is significant only for limited users.

4 Experimental results

4.1 Outline of the experiment

For evaluating a possibility of prediction of information propagation, the traces of information circulation and subjective impressions of information, e.g., interest and importance, are should be investigated simultaneously. Since collecting logs of the communication systems such as

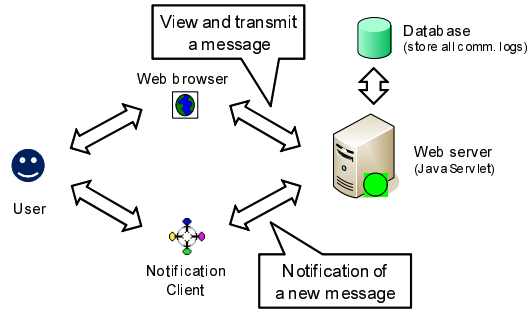

Figure 3. Overview of the experimental system

SNS or BBS are insufficient to evaluate both of them, an information propagation support system that is based on word-of-mouth communication is implemented, and the experiment was carried out by real users.

4.1.1 Experimental system

In order to collect the traces of information circulation, the system must gather contents of information and comments for the information, which propagates among users. Note that comments must be made at the time when a user forwards it because the latest subjective impression affects the user's activity, so the system must provide the user both of contents of information and a questionnaire form for it simultaneously.

Therefore, the experimental system is implemented as a general web application based on Java servlet and MySQL database, as shown in Figure 3. When a user sends a message to others, the user is only forced to login the system, to write contents of the message, and to select target users from the list of participated users. Figure 4 depicts an example screenshot of the system. Consequently, the database of the system stores all of the contents and target users.

In contrast, when a user receives a message from other users, the user is lead to the system by notifying from the client application as shown in Figure 5. The user reads contents of the message, and judges the message is worth to be forwarded to other users. When the user wants to forward it, the user selects target users from the list of participated users. In addition, the system requests the user to answer questionnaires for the message even if the user does not forward it.

As a result, the traces of information circulation and subjective evaluations by the users can be obtained through the database.

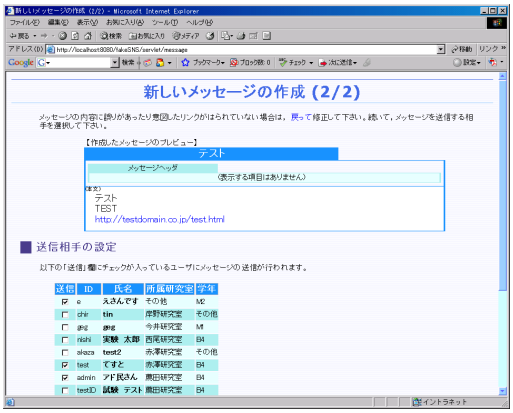

Figure 4. Screenshot of the server when a user sends a message

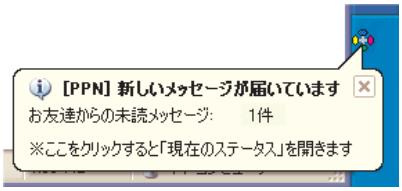

Figure 5. Screenshot of the notification client

4.1.2 Procedure of the experiment

The experiment was performed by 29 students. The participants freely send a message to others, and forwards it when the receiver thinks it is worth to be shared. Questionnaires that are filled by the receivers are organized as follows.

1. Accuracy
   Do you think the message is accurate or not? (Yes / Maybe / Doubtful / No)

2. Interest
   Do you think the message is interesting for you? (5 to 1; 5 indicates very interesting)

3. Importance
   Do you think the message is important for you? (5 to 1; 5 indicates very important)

4.1.3 Overview of the experimental data

The number of messages sent during the period of the experiment is 96. Distribution of the number of users who receive messages and the number of messages corresponds to it is illustrated as Figure 6. Almost of the messages does not propagate, but some messages propagates among almost all of the users.

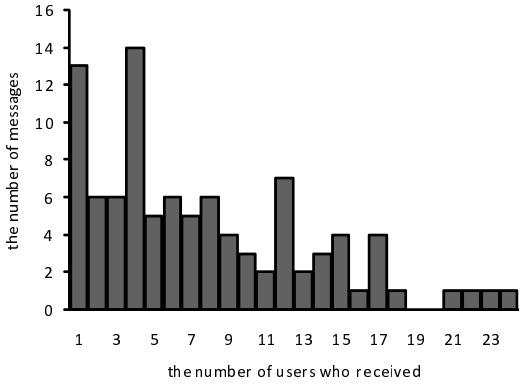

Figure 6. Distribution of the number of users who received the messages

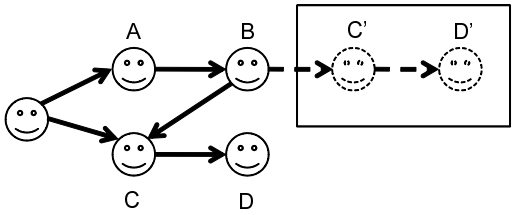

Figure 7. Construction of a propagation tree when a certain user receives the same message multiple times

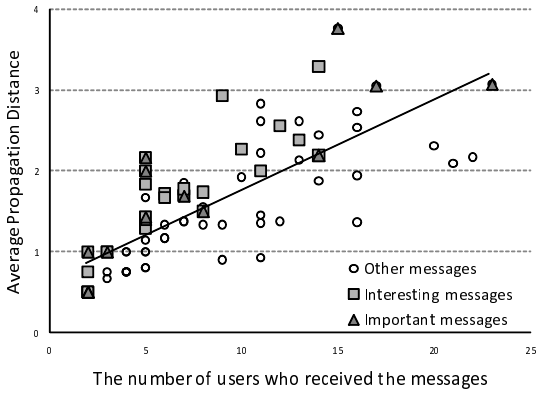

Figure 8. Correlation between the number of users who received the messages and average propagation distance

4.2 Evaluations

4.2.1 Outline of the evaluations

The propagation trees of all the messages are constructed based on the propagation forms of them. The value of average propagation distance and the number of receivers are evaluated based on the trees. Moreover, subjective evaluations are obtained by analyzing the results of questionnaires. Possibilities of prediction of information propagation based on the average propagation distance are evaluated by investigating the correlation between the forms and the questionnaires.

Note that a user selects target users freely so that the user cannot identify who receives the same message. Thus, some user may receive the same message multiple times in this system as shown in Figure 7. When constructing a propagation tree, such a user and the downstream users are duplicated to all of the possible nodes. For instance, in case of user B forwards the message to user C in the figure, the propagation tree is constructed with user C' and D' and the value of average propagation distance is calculated with them. However, the number of users who receive the message is only counted the real receivers, i.e., C' and D' is ignored.

4.2.2 Correlation between the average propagation distance and subjective evaluations

The value of APD is affected by the number of receives, so the distribution of the number of receives and the value of APD is evaluated and illustrated as Figure 8. Meanwhile, most interesting or important 20 messages are selected by the questionnaires, which indicate significant messages by subjective evaluations.

The value of APD tends to increase as the number of receivers increases, but the value of APD is diverse at the same number of receivers. As described in Section 3.3, this difference may indicate the difference of nature of information.

To compare the difference, all the messages are classified into 2 groups which indicates the value of APD is larger or smaller than the linear regression form $y = 0.115x + 0.6096$, where y indicates the value of average propagation distance and x indicates the number of receivers.

Consequently, although the number of receivers does not relate to the interest or importance of the messages, almost of the interesting and important messages are belong to the larger APD group. Thus, interesting or important information, which can be considered to have a potential to propagate among many users, may be identified by the value of APD.

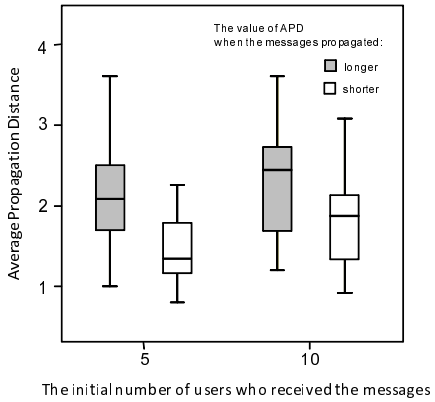

Figure 9. Distribution of APD when the messages propagated until a specific number of users

4.2.3 Differences of the beginning of information propagation

In order to detect information that is possible to propagate, differences between significant and ordinary information at the beginning of propagation should be investigated. Thus, first, all the messages are classified into 2 groups by the value of APD is larger than the linear regression form which is described in Section 4.2.2. Secondly, the messages that propagate more than 5 or 10 users are selected. Thirdly, the value of APD is calculated only for first 5 or 10 receivers of each message. Lastly, distribution of the value of APD is illustrated as a box plot in Figure 9.

As shown in the figure, the value of APD of the messages which belong to the larger APD group tend to be larger than the messages which belong to shorter group at the beginning of propagation, which is verified by t-test. Therefore, significant information's value of APD, which has a potential to propagate, tends to be larger from the beginning.

4.3 Discussions

From the results of evaluations, word-of-mouth communication forms may indicate significance of information. Since such information is considered to propagate, detecting propagation forms of information will enable us to predict undesired information propagation. Although the forms must be obtained by collecting the traces of circulation of information, subjective impressions of the users are not required to predict.

5 Conclusion

In this paper, the average propagation distance was proposed to predict information propagation on the Internet, which can be obtained by word-of-mouth communication. By performing the experiment with real users, the tendency that the average value of propagation distance of important or interesting information is larger than ordinary information was revealed. Therefore, information propagation is expected to be predicated by examining word-of-mouth communication forms.

As a future work, a further experiment is required for validating the results obtained in this paper. In addition, how to apply to the real world such as blogs should be addressed.

Acknowledgement

This work was supported in part by the Research and Development Program of Ubiquitous Network Authentication and Agent (2003), The Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT), in Japan.

References

[1] H. Nanba, N. Okuda, and M. Okamura. Extraction and visualization of trend information from newspaper articles and blogs. In *Proceedings of the 6th NTCIR Workshop*, 2007.

[2] TheWebWatcher. http://www.thewebwatcher.com/.

[3] G. W. Allport and L. J. Postman. *The psychology of rumor*. Holt, New York, 1947.

[4] R. L. Rosnow, J. L. Esposito, and L. Gibney. *Factors influencing rumor spreading: Replication and extension*. Language and Communication, 1988.

[5] C. J. Walker and B. Blaine. *The virulence of dread rumors: A field experiment*, volume 11. Language and Communication, 1991.

[6] S. Takeuchi, J. Kamahara, I. Saeki, S. Teraoka, R. Harada, S. Shimojo, and H. Miyahara. Evaluation of the propagation of information based on the results of the experiments using cellular phones. In *Proceedings of the 13th Data Engineering Workshop*, 2002. (In Japanese).