Modeling of Information Sharing on the Business Social Media

Fujio TORIUMI *, Takashi OKADA **, Shuichiro YAMAMOTO ***, and Kenichiro ISHII ***

Abstract: Social media, which support communication on the Internet, continue to spread rapidly. SNSs and micro-blogs are typical services. This research proposes methods for efficient information sharing through social media, especially intra-firm systems for information sharing of research or expertise. In this research, a simple model of social media is provided, whose users share information by directly referring to other users. The authors focused on a particular intra-firm social media and analyzed the usage of its communities. By applying a clustering method, they are classified into six groups and analyzed their characteristics and occupation rate. Finally, the authors proposed a method that supports information sharing and evaluated it by agent-based simulation using our proposed model. The authors clarified that not only encouraging users to post articles but also giving them an opportunity to refer to the existing community’s history is effective for information sharing.

Key Words: social network services, information sharing, agent-based simulation, social media.

1. Introduction

As the steady growth of new network communication tools continues, the expansion of social media such as social networks services (SNS), and blogs are becoming a social phenomenon impacting societies all over the world. For example, Facebook¹, the largest SNS, has more than 500 million active users [1]. Also, twitter², micro-blog site, has more than 170 million users [2]. Social media have also received attention from many companies, because they are expected to be a good space for information sharing. Therefore, in this simulation, a SNS model to analyze its influence on information sharing is proposed.

In general, the purpose of social media is not always information sharing. Thus, a model of a company SNS that shares information for product development is proposed. This research proposed methods for efficient information sharing through social media, especially intra-firm systems used for information sharing of research or expertise.

The authors analyze a social medium managed in an enterprise to clarify how social media are used in real business settings to propose a model of social media. Simulations are performed with the social media model to clarify its effective features to realize efficient information sharing.

2. Related Studies

Previous work analyzed the relation between knowledge sharing in organizations and weak ties [3]. In this study, human relationships which are called weak ties [4] in sociology, are more effective for knowledge sharing than such weak connections as those between departments.

Other work also analyzed the influence of the social networks of workers on knowledge sharing [5], concluding that both ties and organization structures influence knowledge sharing. Strong ties impact knowledge sharing.

On the other hand, there have been many previous studies of online social media. Adamic et al.[6], for example, studied a university SNS called Nexus and analyzed its structure and the attributes and personalities of its users. Yuta et al.[7], who investigated the network structure of Mixi, discovered a gap in its community-size distribution that is not observed in real social networks and developed a simple model to account for this feature. Ahn et al.[8] compared the structures of three online SNSs, (Cyworld, MySpace, and orcut), each with more than 10 million users. Twitter has also been analyzed [9],[10].

Many researches have addressed modeling and simulations of knowledge sharing or emergent innovations. N. Gilbert et al. modeled product development in the organizations [11] and the structures of academic science [12]. Y. Fujita et al. studied the roles of innovators in information distribution using simulations [13]. Other studies modeled SNSs [14],[15] to simulate human behaviors.

3. Proposal for Social Media Model

3.1 Overview

In this paper, the authors modeled a SNS called the “Chie-no-Wa SNS” managed by the NTT Information Sharing Platform Laboratories. The NTT group utilizes a community site named “Chie-no-Wa,” which can be translated as knowledge harmony, through its subsidiary companies. Chie-no-Wa SNS is one subsystem of Chie-no-Wa. This SNS was developed to share information. Communities and forums are its main features. Therefore, the authors will model it as information sharing sites using the community system using multi-agent based modeling to model the social media.

3.2 Social Media Model

3.2.1 Agent

We define N agents that share the information in an SNS as
Each element \( e_{ij} \) takes binary values \([0, 1]\). All agents have specific objective functions calculated from the information they possess. By repeating information sharing, all agents maximize their objective function values. Compared to product development, this process resembles information collection to find and develop suitable products. Therefore, agents affiliated with the same groups have the same objective functions. The details of objective function will be explained in Section 3.3.

Our model has two types of agents. “Information provider agents” provide information to other agents. “Information browser agents” rarely provide information. The number of information provide agents are decided by the analysis results of a real social network service shown in Section 4.

3.2.2 Community

A community is a system to make groups in a SNS to find other users who have the same interests. All users can write information to the Bulletin board system in communities and read the information written by other users of the Bulletin board system. To represent this community feature, we modeled it as an information repository. The agents in the model participated in several communities. All agents can browse the accumulated information provided by other agents from the communities in which they participated. From the above procedure, all agents maximize their objective function value. The following are the agents behaviors for the community:

- Information provision
- Information browsing

The details of all behaviors are explained in Section 3.4. The outline of information sharing by community is shown in Fig. 1.

3.3 Objective Function using NK-Landscape Model

In our proposed model, all agents participating in the social media have their own objectives. The quality of information to achieve such objectives is represented by an objective function. A higher objective function value represents a high quality of information to develop feasible products. The objective function is modeled using a NK-Landscape model [16]. When we assume information \( E_i \) as one genotype, the fitness of the gene is assumed to match the objective function value. In the NK-Landscape model, the fitness of the genes is evaluated by the value of each locus and its neighbors. The fitness of gene \( E_i \) in the NK-Landscape is evaluated by the following equation:

\[
f_i(E_i) = \frac{1}{N} \sum_{j} f_j(e_{ij}, z_{i1}^{(j)}, \ldots, z_{ik}^{(j)}),
\]

where each locus \( e_{ij} \) takes 0 and 1, and \( z_{ik}^{(j)} \) represents the loci which have relationships with \( e_{ij} \). The fitness of a locus is derived from its value and the related loci. Related loci \( z_{i1}^{(j)}, \ldots, z_{ik}^{(j)} \) are the \( K \)s that surround the locus of loci \( e_{ij} \). When \( j = 3, K = 2 \), the following elements are used to calculate the fitness of \( e_{ij} \):

\[
\{e_{ij}, z_{i1}^{(j)}, \ldots, z_{ik}^{(j)}\} = \{e_{i2}, e_{i3}, e_{i4}\}.
\]

Considering that each locus \( e_i \) takes a binary value, array \( \{e_{ij}, z_{i1}^{(j)}, \ldots, z_{ik}^{(j)}\} \) becomes a \( K + 1 \) length bit array.

In the NK-Landscape model, we prepare fitness arrays for the bit arrays. A sample of the fitness array for \( K = 2 \) is shown in Table 1. When the bit array for locus \( e_{ij} \) is \( \{e_{ij}, e_{i(j−1)}, e_{i(j+1)}\} = \{010\} \), the fitness of \( e_i \) is calculated from the fitness array as follows:

\[
f_i(e_{ij}, e_{i(j−1)}, e_{i(j+1)}) = f_i(010) = 0.653.
\]

The fitness of the gene is calculated from the average of each locus.

Each locus has its own fitness array generated by \( 2^{K+1} \) random value that takes \([0.0, 1.0]\). Considering that each locus has a different fitness array, to evaluate all of the gene loci, a fitness table comprised of \( N2^{K+1} \) random values is generated beforehand.

3.4 Agent Behavior

3.4.1 Information provision

All agents provide their information to the community at a certain rate. This is a representation of the written information to the BBS in a community in real social media. The following are the information provision procedures. First, we decide whether the provision has occurred in certain steps from the provision rate calculated from the amount of provided information in the last 30 steps. The detail of the information provision is explained in Section 5.1. Next, if the provision is determined to have occurred, the agent who provided the information is selected. Information provider agents are selected first. When the information provider agents are not selected, the other agents are selected randomly. Finally, information \( E_i \), which selected agent \( a_i \) is provided to the community. The information provided by the agent is retained for a certain period. While the information is being retained, the other agents who access the community can browse it.
3.4.2 Information browsing

The agent selects one piece of information from the browsed information, which is the latest $P_t$ information that is provided to the participating community. The tournament selection [17], which selects the best information from a few random bits of selected information, selects the information.

3.4.3 Information learning

The agent learns the selected information. The learning procedure is represented by the idea of crossover in the Genetic Algorithm (GA). A piece of information is assumed to be a gene. By crossovering two pieces of information, the agent generates new information. If the new information has higher fitness, the agent updates his/her information as new information.

4. Community Analysis

4.1 Classification of Community

To realize a realistic simulation, the authors analyzed real social media. By using analysis results, the authors realized realistic parameter settings for the simulation to improve its reliability. Since many types of social media communities exist, the authors classified them by their features and clarified how they are used in social media to design a reliable communication model.

4.2 Feature Vector

We classified communities on a SNS called “Chie-no-Wa SNS,” managed by the NTT Information Sharing Platform Laboratories. To classify the communities, we employed the K-Means classification algorithm [18] and analyzed all types of communities classified by it from the viewpoint of communication through communities.

The following feature vectors are used for classification:

- Number of posted articles: $P_m$
- Number of participating users: $U_b$
- Number of information providers: $U_p$
- Number of articles posted by the most active user: $P_{max}$

The information providers are the users who post articles to the community. The normalized value of each feature vector is used to solve the algorithm.

4.3 Classification Results

The feature vectors of types of communities $C_1, \cdots, C_6$, which were obtained from the classification, are shown in Table 2.

Users in the community belonging to type $C_1$ often post articles. The number of articles is the largest, and almost 25.5% of the users post articles. On the other hand, half of the articles are posted by the most dedicated user: the active user in such communities.

In $C_2$ type communities, many users participated. The rate at which articles are posted by the most active user is smaller than $C_1$. Also, the number of users who posted articles exceeds $C_1$, showing that in the $C_2$ type communities, many users actively communicated with each other. In $C_3$ type communities, there is some communication but less than in the $C_1$ type communities. The rate of information providers is only 15.7% in such communities. In $C_4$ type communities, the average number of participants is larger than $C_1$ type communities. However, no articles were posted to the community; in other words, the utilization rate of such communities is extremely small. $C_{end}$ and $C_6$ types of communities have fewer participants and fewer posted articles. Thus, these communities are not being used.

In the next section, we will simulate the effect of information sharing on social media using classification results.

5. Simulation of Information Sharing by Community

5.1 Simulation Aim and Setting

The aim of this simulation is to represent the information sharing by communities on a social medium.

To realize a realistic simulation, we used the result of community analysis that was explained in Section 4 as parameter settings. Every community represented in the simulation has the following parameters:

- Number of posted articles $P_m$
- Number of participating users $U_b$
- Number of information providers $U_p$
- Rate of articles posted by the most active user $P_{max}$

The number of posted articles $P_m$ represents the number of articles posted to the communication in 30 steps. In all steps, the ratio at which articles are posted to the community is $P_m/30$. The number of participating users $U_b$ represents how many agents participated in the community. The number of information providers $U_p$ represents the number of users who posted articles to the community. Such users are decided in the preparation phase, which never changes during the simulation. On the other hand, all agents can browse the information from the community.

The rate of articles of the most committed user $P_{max}$ represents how many articles are posted by the active user. When $P_{max}$ is large, the community is led by the active user. On the other hand, when $P_{max}$ is small, debates often occur in which many users participated. Using the above parameters, we modeled the six types of communities (Table 2). The types of communities are randomly set to $C_1 - C_6$ based on the occupancy. The other simulation settings are shown in Table 3.

5.2 Evaluation

The simulation result was evaluated by the fitness averages of the information calculated from the objective function. When the fitness increases, the mechanism of information sharing by community is effective for product development.

5.3 Base Simulation Results

First, we calculated the base simulation that has the above parameters. The simulation result is shown in Fig. 2. The x-axis
Table 3 Simulation settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agents $M$</td>
<td>1000</td>
</tr>
<tr>
<td>Communities $C$</td>
<td>150</td>
</tr>
<tr>
<td>Browsable information $P$</td>
<td>10</td>
</tr>
<tr>
<td>Objective function $L_a$</td>
<td>5</td>
</tr>
<tr>
<td>Length of information $N$</td>
<td>20</td>
</tr>
<tr>
<td>Related locus $K$</td>
<td>5</td>
</tr>
<tr>
<td>Tournament size $S_t$</td>
<td>10</td>
</tr>
<tr>
<td>Simulations</td>
<td>100</td>
</tr>
</tbody>
</table>

Fig. 2 Base simulation results.

shows the simulation steps, and the y-axis shows the average objective function values of all agents. The result is the average of 100 simulations.

The simulation result clearly shows that for the first 10 steps, the average value of the objective function increased rapidly. However, after 20 steps passed, the increment rate of the objective function value decreased. Although the maximum objective function value is $1.0$, the final objective function value is insufficient. Therefore, in the next section, we seek a more effective method to support information sharing.

6. Simulation of Efficient Method for Information Sharing

6.1 Simulation Aim and Settings

The simulation aim is to find a method to support efficient information sharing. The simulation parameters are set to the same simulation explained in Section 5.

In this simulation, the following two methods are proposed to support information sharing. The first method is large community, and the second is community recommendation. The validity of all methods are checked by simulation.

The authors proposed these two methods for the following reasons. The social media we modeled was developed to realize information sharing in companies. In such cases, since methods that require drastic system changes are problematic, we chose methods that only require slight change for systems.

6.2 Information Sharing Support Methods

6.2.1 Large community

In this method, a large community in which all users participate is created in the social media model. In this community, an article is posted by following all the steps.

This model represents the introduction of a company-wide community. Such a community often encourages inter-department communications. Also, with large communities, the amount of information that agents can browse is expected to increase.

6.2.2 Community recommendations

In this method, a community is recommended to all agents at a certain rate (10% here). When the agent recommends a community, participation increases. This model represents the community recommendations from other users. For example, when a department requires new technologies to develop products, its members recommend their own communities to other company members who have knowledge of the technologies to share the information of new technologies.

The following two types of recommendations are simulated here:

- To agents with the same objective function
- Randomly to agents

6.3 Simulation Result and Discussion

The simulation result is shown in Fig. 3. The x-axis shows the simulation steps, and the y-axis shows the averages of the objective function values of all agents. Additionally, the simulation calculated in Section 5 is shown as a base line result.

From Fig. 3 all methods that support information sharing effectively improve the average fitness of the information of all agents. In other words, both methods are effective for information sharing support. Different characteristics are confirmed in each change in the fitness line.

The large community result shows that fitness increased drastically in the first few steps, but became saturated in the early periods because all agents were given a chance to find various users, which improved the information fitness. However, the information that can be found in large communities does not always elevate the information fitness of agents. On the other hand, the community recommendations realized smooth but not saturated improvement of information fitness. Finally, the fitness of the community recommendations became larger than that of the large communities.

From the difference between the results of recommendations to an agent with the same objective function and random recommendations to an agent, objective function has a greater affect than information sharing. However, random recommendations are also effective. Thus, recommendations seem to be an extremely effective method to support information sharing. In the recommendation methods, the fitness of the agent information improved despite no increments of article posts because the agents have more opportunities to find information that was already provided to the community. Such additional information
provides chances to elevate the fitness of the information. From the simulation result, we predict that users must contact much information to raise its value. Therefore, introducing a new system that recommends other communities in which users are not participating effectively supports information sharing. For example, one implementation might be a system that finds other users with knowledge about the topics when the community discussions become deadlocked.

7. Conclusion

In this research, the authors provided a simple model of social media, whose users are sharing their information by a community. To realize realistic parameters for the model, the authors focused on a particular intra-firm social medium and analyzed the usage of its communities. By applying a K-Means clustering method, we classified them into six groups and analyzed their characteristics and percentages.

Finally, the authors proposed methods that support information sharing and evaluated them by an agent-based simulation using our proposed model. The authors clarified that not only encouraging users to post articles but also giving them the opportunity to refer to the existing community’s history is effective for information sharing.

Future work includes modeling other social media and implementing our proposed methods. To enhance the cogency of each model including learning process is one of the most important future works.

Acknowledgments

This research was conducted under the collaboration scheme between Nagoya University and NTT. The authors would like to thank the scheme and people who supported it.

References


He received Ph.D. from Tokyo Institute of Technology in 2004. He joined Nagoya University as Assistant Professor in 2004. And he moved the University of Tokyo as Associate Professor. Recently, he is engaged in research of social media analysis and agent-based simulation.

Takashi Okada

He received B. Eng. and M. Eng. degree from the University of Tokyo in 1987 and 1989 respectively. Since 1989, he is working for NTT (Nippon Telegraph and Telephone Corp.) laboratories. Recently, he is engaged in practical research of knowledge management and implementation of the knowledge management process based on new generation communication system in the NTT group.

Shuichiro Yamamoto


Kenichiro Ishii

He received M. Eng. degree from the University of Tokyo in 1974 and he joined NTT. He received D.Eng. degree from the University of Tokyo in 1990. And he moved Nagoya University in 2003 as Professor.