Proc. of IFIP 4th International Conference on Entertainment Computing (ICEC 2005), Sep. 2005, Sanda, Japan, published in Lecture Notes in Computer Science, Fumio Kishino et al. (Eds.), vol. 3711, pp. 148-154.

Keyword Discovery by Measuring Influence Rates on Bulletin Board Services

Kohei Tsuda and Ruck Thawonmas*

Intelligent Computer Entertainment Laboratory,
Department of Human and Computer Intelligence, Ritsumeikan University,
Kusatsu, Shiga 525-8577, Japan
ruck@ci.ritsumei.ac.jp

Abstract. In this paper, we focus on relations between comments on Tree-style Bulletin Board Services (BBSs), and propose a method for discovering keywords by measuring influence rates thereon. Our method is based on an extended model of Influence Diffusion Model (IDM) proposed by N. Matsumura et al. in 2002, where they discussed the influence diffusion of a term in a comment to all succeeding comments that include that term and reply to that comment. Here we additionally consider the influence diffusion of a term over comments that include that term and all reply to a same comment, as well as the influence diffusion of a term over nearby comments that include that term, regardless of their reply relation. Evaluation results using Tree-style BBS data related to Massively Multiplayer Online Games (MMOGs) show that the proposed method has higher precision and recall rates than IDM and a classical method based on term frequencies. As a result, keywords discovered by the proposed method can be effectively used by MMOG publishers for incorporating users' needs into game contents.

Keywords: Keyword Discovery, Tree-style BBSs, Comments, MMOGs.

1 Introduction

Online contents such as Bulletin Board Services (BBSs) have recently gained a lot of attention as new tools for market analysis [1]. In Massively Multiplayer Online Games (MMOGs), their contents must be accordingly updated after the first release in order to retain users [2]. It is therefore important to grasp user demands, and for this task BBSs outside of the game are a good candidate. With the increasing number of documents in such BBSs, automatically discovering keywords is a very challenging and important issue.

We focus on relations between comments on Tree-style BBSs, where the reply relation of comments (who replies to whom) is clear. For keyword discovery

* The author has been supported in part by Ritsumeikan University's **Kyoto Art and Entertainment Innovation Research**, a project of the 21^{st} Century Center of Excellence Program funded by Ministry of Education, Culture, Sports, Science and Technology, Japan; and by Grant-in-Aid for Scientific Research (C), Number 16500091, the Japan Society for Promotion of Science.

F. Kishino et al. (Eds.): ICEC 2005, LNCS 3711, pp. 148–154, 2005.

on such BBSs, we extend Influence Diffusion Model (IDM) [3], [4] and propose our method based on it. The main concept behind our method is that an influential term in a comment should be also used in succeeding comments. In our method, three types of influence diffusion of a term over comments that include that term are considered, namely, the influence diffusion from a comment to all succeeding comments that reply to that comment (originally discussed in IDM), the influence diffusion over comments that all reply to a same comment, and the influence diffusion over nearby comments, regardless of their reply relation. In our case study using Tree-style BBS data related to MMOGs, the proposed method, IDM, and a classical method based on term frequencies are compared. Thereby the superiority of the proposed method over the others is confirmed in terms of both the precision rate and the recall rate.

2 Measuring of Influence Rates

In this section, we describe the definition of comment relations and our algorithm for measuring influence rates. On BBSs, communication is done by exchanging comments, via posting a new comment and its reply comments. The influence of a term is diffused from a comment in which the term resides to succeeding comments that include also that term.

2.1 Comment Relations

We consider that there are three types of comment relations, via which the influence of a term is diffused over comments that include also that term, as follows:

Comment Chain. (Fig. 1(a)) that shows the influence diffusion of a term from a comment, say comment **X**, to all succeeding comments that reply to comment **X**;

Parallel Chain. (Fig. 1(b)) that shows the influence diffusion of a term from a comment that replies to a preceding comment, say comment **Y**, to all succeeding comments that also reply to comment **Y**; where terms residing in Comment Chain are excluded from consideration;

Cross Chain. (Fig. 1(c)) that shows the influence diffusion of a term from a comment, say comment **Z**, to up to its α succeeding comments; where terms residing in either Comment Chain or Parallel Chain are excluded from consideration;

where in Fig. 1 comment numbers indicate the order where the comments are posted.

2.2 The Proposed Algorithm

Treating each term equally, we define the influence rate of term t on comment k as follows:

$$i_{t,k} = \frac{1}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|}, \tag{1}$$

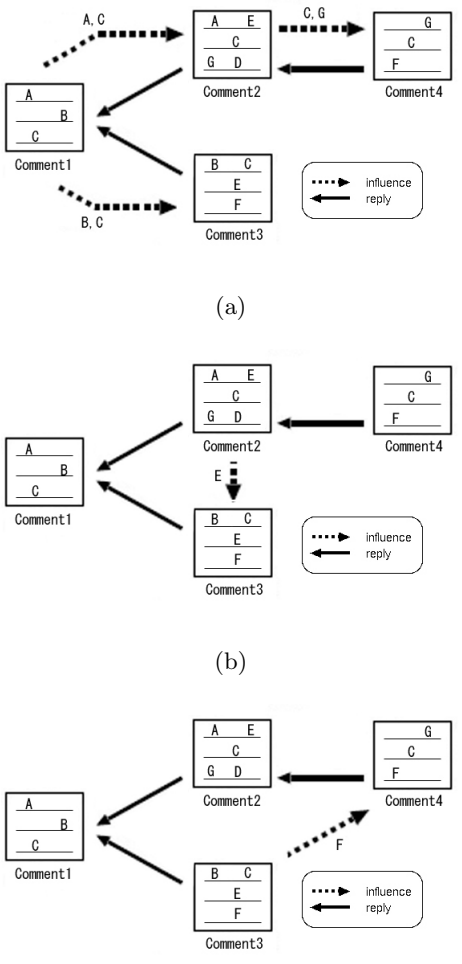

Fig. 1. Examples of (a) Comment Chain, (b) Parallel Chain, and (c) Cross Chain

where $w_{c,k}$, $w_{p,k}$, and $w_{x,k}$ are the set of terms whose influence on comment k is via Comment Chain, Parallel Chain, and Cross Chain, respectively.

In practice, the above three chain types should have different effects in calculation of influences. Our heuristic is as follows:

Effect of Comment Chain $\geq$ *Effect of Parallel Chain* $\geq$ *Effect of Cross Chain*

We further analyzed a targeted BBS in our case study discussed in Section 3 and found that in a given comment, the number of terms whose influence is via a particular chain type basically fits the relation in the above heuristic. We

therefore define the influence rate of term t on comment k via Comment Chain, Parallel Chain, and Cross Chain, respectively, as follows:

$$i_{c,t,k} = i_{t,k} \cdot \frac{|w_{c,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \tag{2}$$

$$i_{p,t,k} = i_{t,k} \cdot \frac{|w_{p,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \tag{3}$$

$$i_{x,t,k} = i_{t,k} \cdot \frac{|w_{x,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \tag{4}$$

Due to the definition of each chain type, the influence of a term on a given comment is via either of the three chain types. As a result, the total influence rate of term k diffused from comment 1 to comment n can be given below as:

$$I_{t,n} = \sum_{k=1}^{n} j_{t,k}, \tag{5}$$

where $j_{t,k}$ is defined as

$$j_{t,k} = \begin{cases} i_{c,t,k} & \text{if via Comment Chain} \\ i_{p,t,k} & \text{if via Parallel Chain} \\ i_{x,t,k} & \text{if via Cross Chain} \end{cases} \tag{6}$$

Now we are ready to give the algorithm for measuring influence rates. Our algorithm is as follows:

1. Decide the first comment and the last comment for measuring influence rates in a given BBS, and then move to the first comment
2. Move to the next succeeding comment
3. Select all terms in the current comment whose influence is via Comment Chain
4. Select all remaining terms in the current comment whose influence is via Parallel Chain
5. Select all remaining terms the current comment whose influence is via Cross Chain
6. Calculate the influence rate of each selected term using (5)
7. Repeat 2 to 6 if the current comment is not the last one.

3 Case Study

We analyzed 1697 comments in Japanese that were posted on Yahoo!BBS [5] (Fig. 2) during June 22, 2002 and August 15, 2004. The main topic of this BBS is on MMOG systems, especially on new ideas and dissatisfactions of users. From

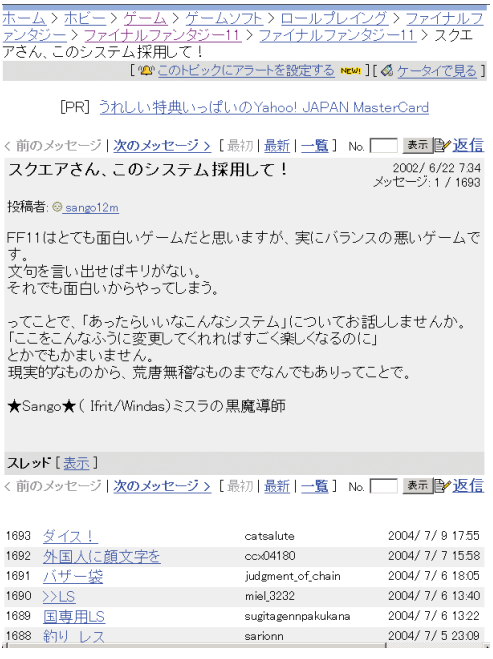

Fig. 2. Yahoo!BBS used in our case study

these comments, 4906 terms (only nouns) were extracted. We compare the result of the proposed method with that of IDM and that of a classical method using Term Frequencies (TF)[1].

3.1 Influential Terms

The top 10 influential terms from each method, with $\alpha = 2$ for our method, are listed in Table 1.

3.2 Precision and Recall Rates

We asked 10 human subjects, 5 veteran players playing MMOGs more than 100 hours monthly and 5 novice players playing less, to thoroughly read all comments. We then separated the subjects into two groups, i.e., a group of the veteran players and a group of the novice players. For each group, its members were asked to individually select 100 important terms, and the terms selected in common by all members were considered influential terms of that group. By this, 6 influential terms were decided by the former group, and 5 influential terms by the latter group. Because there was one influential term decided in common

[1] In TF, terms are simply ranked according to their occurrence frequency in all comments.

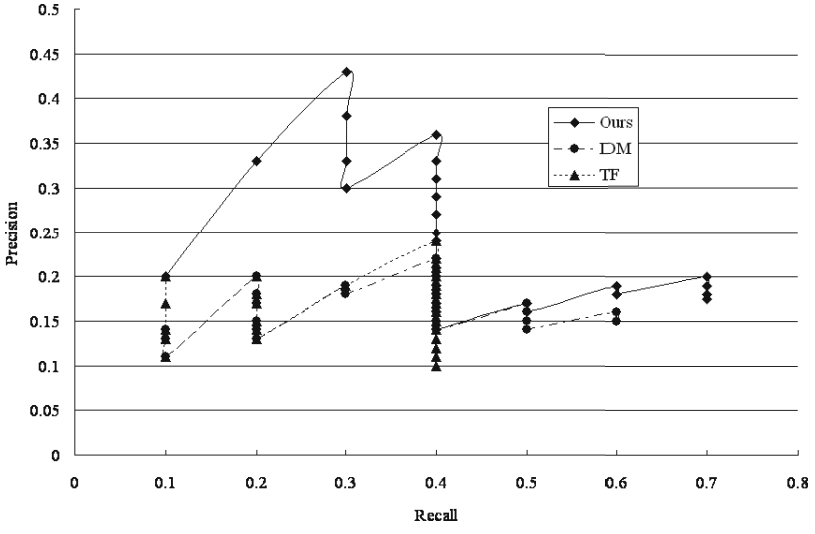

Fig. 3. Precision/Recall curve of each method

Table 1. Influential terms from each method

Ranking	Ours	IDM	TF
1	System	Human	System
2	Human	Level	Human
3	Level	Magic	Level
4	Game	Warrior	Game
5	Balance	Experience	Magic
6	Item	Equipment	Charm
7	Equipment	Red (magic)	Warrior
8	Experience	Attack	Experience
9	Self	White (magic)	Equipment
10	Magic	Charm	Black (magic)

Table 2. Influential terms decided by human subjects

	Term		
Veteran	Item	Cure	Equipment
	Individual	Quest	Solo
Novice	Balance	Charm	
	Combination		Casino

by both groups, the total number of influential terms decided by the human subjects is 10, and they are listed in Table 2.

Figure 3 shows the precision rates over the recall rates of the proposed method, TF, and IDM, where the influential terms decided by the human subjects were considered correct ones. This figure shows that the proposed method is of higher performance than the others for all ranges.

4 Conclusion

In this paper, we proposed new concepts on comment relations on BBSs and the method based on these concepts for measuring influence rates. Like IDM, our method is yet another formalization to understanding of diffusion of influential terms on internet-mediated communication, which has recently attracted attention from researchers on online communities. We have shown using real BBS data related to MMOGs that the proposed method outperforms the existing methods, IDM and TF. MMOG publishers thus can use the proposed method for discovering users' needs and incorporate them into game contents. In our future work, we plan to modify and apply our method to other types of entertainment-oriented text-based communities, such as blogs, chat rooms, and social networking services.

References

1. Hanson, W.: Principles of Internet Marketing. South-Western Pub. (1999)
2. Alexander, T. (ed): Massively Multiplayer Game Development. Charles River Media, Inc., Massachusetts (2003)
3. Matsumura, N., Ohsawa, Y., Ishizuka, M.: Influence Diffusion Model in Text-Based Communication. Proc. Int'l World Wide Web Conf. (WWW02), Hawaii (2002)
4. Matsumura, N., Ohsawa, Y., Ishizuka, M.: Influence Diffusion Model in Text-Based Communication. Journal of the Japanese Society for Artificial Intelligence, Vol. 13, No. 3, 259–267 (2002) (in Japanese)
5. Yahoo!BBS. http://messages.yahoo.co.jp/bbs