Proc. IFIP 5th International Conference on Entertainment Computing (ICEC 2006), Sep. 2006, Cambridge, UK, published in Lecture Notes in Computer Science, Richard Harper et al. (Eds.), vol. 4161, pp. 286-291.

Discovery of Online Game User Relationship Based on Co-occurrence of Words

Ruck Thawonmas*, Yuki Konno, and Kohei Tsuda

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

Abstract. Understanding of users is an important key to keep an online game fascinating to them. In this paper, we discuss a heuristic that associates relationship between users with co-occurrence of words uttered by them in a given period. An experiment is conducted using Ragnarok Online logs in the client-side of one human subject, which confirms the effectiveness of the proposed method based on this heuristic.

1 Introduction

The MMOG (Massively Multiplayer Online Game) industry is one of the fastest growing industries. The Themis Group estimated in one of their white papers [1] that worldwide revenues of MMOGs will rise from 1.30 Billion USD in 2004 to 4.10 Billion USD in 2008, and to 9 Billion USD in 2014.

In MMOGs, social relationship among players is naturally formed. For MMOG companies, it is important to discover such relationship in order to maintain a good community. This is because being in a good community usually makes players addicted to the game.

Communications in MMOGs are mainly based on chat. In the research field called social network analysis, the strength of relationship between members of a group can be analyzed through their communications [2]. This is usually done by asking the question "who do you communicate with?" to group members.

Automatic approaches also exist for identifying conversation chatters. For example, PieSpy [3] uses heuristics such as addresses mentioned in messages as well as consecutive messages from only two users. The performance analysis of PieSpy, however, has not been studied and reported in the literature.

Another approach proposed in [4] is based on the assumption that users who utter messages in a same interval are interacting with one another. In this approach, time is divided into multiple slots, and users are grouped into clusters

* The author was 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.

R. Harper, M. Rauterberg, M. Combetto (Eds.): ICEC 2006, LNCS 4161, pp. 286–291, 2006.

Fig. 1. A screen shot of an avatar chat where word co-occurrence is seen

based on input features extracted by singular value decomposition (SVD) from their co-occurrence in each slot. Due to not considering discussion topics, the approach has arguably low performance in discovering conversation chatter pairs, as shown in [4], when multiple discussions are on-going concurrently by many chatter groups.

2 User Relationship Based on Co-occurrence of Words

The aim of our study is to examine a new heuristic for use in conjunction with the heuristics mentioned above. We focus on discussion topics. Our heuristic is that if chatters utter messages that contain common words in a particular period, they are participating in a same discussion and thus are communicating to one another. In other words, this heuristic is based on co-occurrence of words uttered by different users in a given period. Figure 1 shows a screen shot of an avatar chat, where there exists one common word, those surrounded by the box, muttered by multiple chatters who are talking to each other.

To implement our heuristic, a list data structure is employed for each user. Each list contains up to W different meaningful words uttered by its user. The definition of meaningful words is given in the next section; henceforth, we simply use word(s) for meaningful word(s), unless stated otherwise. In addition, words whose span $\geq \tau$ are removed from the list. If a message is uttered by a user, each word in the message is first tagged with a present time stamp and is then operated as follows:

- If the user's list does not already contain it and is not full, the new word of interest will be placed in the list.
- If the user's list does not already contain it but is full, the word with the oldest time stamp in the list will be replaced by the new word of interest.
- If the user's list already contains it, the time stamp of the word of interest in the list will be updated.

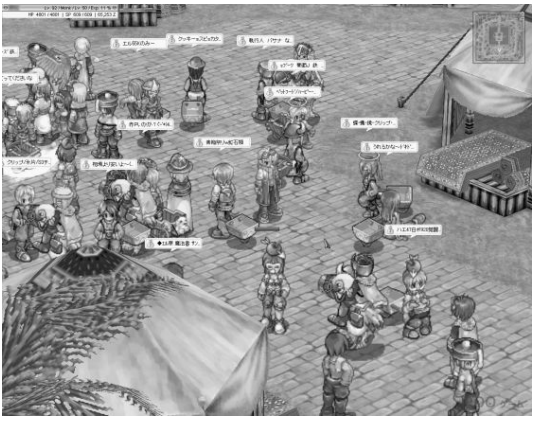

Fig. 2. A screen shot of Ragnarok Online used in the experiment

Now we describe how to compute user relationship based on our heuristic. Let t, $S(t)$, and $L(t)$ denote the message number, the sender of message t, and one of the listeners of this message, respectively. In addition, let $\omega_{L(t)}$ and γ_t represent the set of words in the word list of $L(t)$ and the set of words in message t, respectively. The increase in the relationship of $L(t)$ toward $S(t)$ due to message t, $\delta_{L(t),S(t)}$, is defined as follows:

$$\delta_{L(t),S(t)} = \frac{\mid \omega_{L(t)} \cap \gamma_t \mid}{W} \tag{1}$$

For a period of n messages, the total increase in the relationship of user A toward user B, $T_{A,B}$, is computed as follows:

$$T_{A,B} = \sum_{t=1}^{n} I_{A,B}(t), \tag{2}$$

where $I_{A,B}(t) = \delta_{L(t),S(t)}$ if $A = L(t)$ and $B = S(t)$; otherwise, $I_{A,B}(t) = 0$.

For this period, the mutual relationship between users A and B, $M_{A,B}$ or $M_{B,A}$, is given as follows:

$$M_{A,B} = M_{B,A} = T_{A,B} + T_{B,A} \tag{3}$$

3 Experiment

To examine its effectiveness, we tested the proposed method with chat logs of one subject who is a user of Ragnarok Online (RO) [5], a popular Korean online game in Japan whose screen shot [6] is shown in Fig. 2. These logs were obtained over a period of six months at the subject's PC, and they contained about sixty thousand messages.

There are four chat modes in RO, namely, normal chat, party chat, guild chat, and whisper chat, described as follows:

Table 1. Rankings in the relationship between the subject and his most favorite five user characters by the subject himself, by co-occurrence of words (proposed method), and by the number of messages

No.	subject himself	co-occurrence of words	number of messages
1	A	A	E
2	B	B	A
3	C	-	-
4	D	D	B
5	E	C	-
6	-	E	-
7	-	-	-
8	-	-	-
9	-	-	D
10	-	-	-
11	-	-	-
12	-	-	-
13	-	-	-
14	-	-	-
15	-	-	C

- In the normal chat mode, a user of interest can chat with other users who are within the chat range defined by RO.
- In the party chat mode, a user of interest can chat with other users who have formed a party for a particular mission with him.
- In the guild chat mode, a user of interest can chat with other members of a guild where he belongs.
- In the whisper chat mode, a user of interest can secretly chat with particular users.

In our study, in order to test the generalization ability of the proposed method, we did not distinguish chat messages based on their modes. In addition, we did not take into account whisper-mode messages because sending messages in the whisper chat mode of a user do not remain in client-side logs.

For the proposed method, meaningful words were both noun-type words and unknown-type words that are parsed and categorized by ChaSen [7], a morphological parser for the Japanese language. Unknown-type words are words not registered in the word database of ChaSen. They are important because they represent slang words or RO oriented special worlds. However, unknown-type words consisting of too few or too many characters are usually from looters; as a result, only those with the length between two and twenty characters were considered meaningful words in this experiment[1].

[1] From the authors' experience, most of the unknown-type words with one character are symbols with no meaning, and those with more than twenty characters are usually URL addresses of sites whose contents are not related to the game.

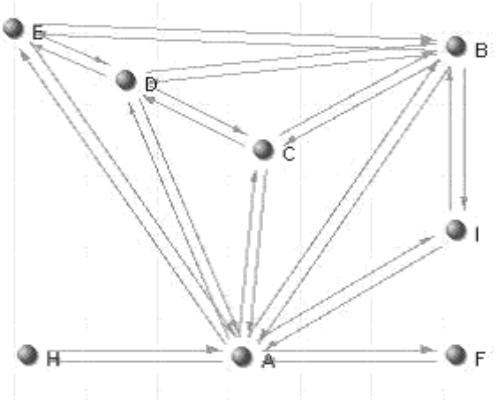

Fig. 3. An example of a social network of game users

The subject was asked to rank the most favorite five other user characters. This ranking is shown in the second column of Table 1, where user character names are represented by aliases. The third column of this table shows the ranking of these five user characters by the proposed method (eqn. 3), where the valid span of each word, τ, and the maximum number of words in the list, W, were set to 10 minutes and 100 words, respectively. In the last column of the table, the ranking by another method is shown that ranked user characters according to the number of their messages in the subject's logs. This table indicates that the proposed method is more accurate than its counterpart.

4 Conclusions and Future Work

We have proposed the method that discovers the relationship between chatters based on co-occurrence of words in their massages. For the RO logs of the subject in the experiment, the proposed method has the result comparable to the result given by the subject himself.

The proposed method is applicable also to chat logs in the server-side, from which social networks of all game users can be constructed, as shown in Fig. 3. This network information can be exploited for obtaining high user satisfaction. One example is that of providing more in-game social activities or events to users with close relationship. Another possible application is that of detecting looters. In addition, if one wants to know more about action behavior of a group of related users, a visualization technique discussed in [8] can be applied to a group of interest.

We do not claim, however, that the proposed method based on word co-occurrence is always better than those in [3] and [4]. In fact, as our future work, we plan to find an effective strategy that combines all of these heuristics for practical use.

References

1. Alexander, K., Bartle, R., Castronova, E., Costikyan, G., Hayter, J., Kurz, T., Manachi, D., Smith, J.: The Themis Report 2004 – Preview. (2004) http://www.themis-group.com/uploads/Themis%20Report%202004%20-%20Preview.pdf
2. McCarty, C.: Social Network Analysis. BEBR Research Report (2003) http://www.bebr.ufl.edu/Articles/SNA_Encyclopedia_Entry.pdf
3. Mutton, P.: Inferring and Visualizing Social Networks on Internet Relay Chat. Proc. Eighth International Conference on Information Visualisation (IV'04) (2004) 35–43
4. Çamtepe C., Krishnamoorthy, M., Yener, B.: A Tool for Internet Chatroom Surveillance. Lecture Notes in Computer Science. Chen, H., Moore, R., Zeng, D., Leavitt, J. (eds.) **3073** (Second Symposium on Intelligence and Security Informatics, ISI 2004) 252–265
5. Ragnarok Online Japan. http://www.ragnarokonline.jp/
6. Ragnarok Online Gallery. http://game.goo.ne.jp/contents/oexgallery/Ragnarok/gallery3.html
7. ChaSen: A morphological parser for the Japanese language. http://chasen.naist.jp/hiki/ChaSen/
8. Thawonmas, R., Hata, K.: Aggregation of Action Symbol Subsequences for Discovery of Online-Game Player Characteristics Using KeyGraph. Lecture Notes in Computer Science. Fumio Kishino et al. (eds.) **3711** (Proc. of IFIP 4th International Conference on Entertainment Computing, ICEC 2005) 126–135