A Method for Conversational Topic Recommendation to Appropriate User Group

Yoshio Ishizawa, Dai Kusui, Hironori Mizuguchi, Yukitaka Kusumura, Yusuke Muraoka
NEC Information and Media Processing Laboratories
8916-47 Takayama-Cho
Ikoma, Nara 630-0101, Japan
{y-ishizawa@bq, kusui@ct, hironori@ab, kusumura@az, y-muraoka@bu}jp.nec.com

ABSTRACT: *As intimacy in human relationships is getting reduced, social problems increase, and a number of people feel cut off. To support making friends and activating communication, we propose a topic recommendation method for conversation to an appropriate user group. Although previous methods have an issue where many of persons may not enjoy conversation by the topic recommendation method, our proposed method can provide the topics which are preferred by more persons for their conversation, and more persons can enjoy their conversation. The method utilizes the similarities in all combination between topic and a user preference. The topic preference vector which is formed by user as the element and the similarity as the value is constructed for each topic, and the topics are clustered by using the topic preference vectors, based on the similarities. Then, the user groups are extracted from the topic clusters, and the corresponding topics are recommended to the user groups. For the proposed method, the effectiveness of user group construction, the completeness of the user preference extraction and the appropriateness of topic selection are investigated by conducting an simple experiment. As the result, we confirmed the user group construction can help to have more enjoyable conversation, and although the user preference extraction was not enough, the topics offered by the proposed method and along with a category are appropriate for the known persons and the unknown persons, respectively.*

Keywords: Social communication, Aging society, Communication Activation, Topic recommendation

Received: 12 July 2011, Revised 14 August 2011, Accepted 17 August 2011

1. Introduction

As the population of senior citizens increases, social problems, which includes lonely-death and elderly person's crimes, receive more attention in Japan. Moreover, the number of relatives and close friends who take care of elderly persons is decreasing, because intimacy in human relationships is getting reduced. The Japanese government reported that elderly person with close friends feels that the life is more fulfilling than without such friends. This suggests that it is important for senior citizens to make close friends to help them find motivation in life.

Although a person would like to talk to others to make friends, it is difficult for the person to enjoy conversation with the partners, especially in case the partners are unknown persons, without any knowledge, such as their histories, preferences, etc. Therefore, it is useful to support their conversation under such situation. One of the ways to support their conversation is to recommend preferable topics. Several topic recommendation systems to enjoy their conversation for the purpose of making friends are proposed in real and virtual spaces[1, 2, 3, 4].

Many of the systems for conversational topic recommendation in real space utilize user profile, such as user preferences which

are taken from questionnaire, utterances, web browsing history etc. The users are grouped by measuring similarities between user preferences, and the system offers appropriate topics which is along with the user preferences for each user group. For example, Fujimoto et al. assumes that the utterances in the user group conversation reflect user preferences for the participants, and developed a system which measures similarity from conceptual relations between user utterances and topic candidates. Finally, the system outputs moderately similar topics[1].

Instead of starting from depending on user preferences, the other systems start from constructing topic clusters which contain similar topics. These systems make topic clusters using similarity between topics. Then, the systems find users who prefer the topics in the cluster, and recommend the topics to the corresponding users.

However, persons have a wide variety of preferences. As the result, many of persons who meet together in one place may not always enjoy their conversation with conversation partners using the topic recommendation systems. The system should recommend topics to as many of persons as possible to enjoy their conversation. This issues come from handling of the topics and the user preferences separately. The topics and the user preferences should be handled together.

In this paper, we propose and evaluate a method for conversational topic recommendation to appropriate user group. This method utilizes topics and user preferences, and measures the similarities between topics and user preferences. Then, topics are clustered by using the similarities, and an appropriate user group which corresponds to each topic cluster is extracted.

The contents of this paper are organized as follows: Section 2 contains a description of basic idea for topic recommendation and the topic recommendation system. Section 3 goes into the details of the simple experiment, and the results are found in Section 4. Section 5 provides the conclusion.

2. Architecture

In this section, the basic idea and the implemented system are described. We assume that it is easy to have their conversation, if participants in conversation know topics which other participants prefer along with their user preferences.

2.1 Basic Idea for Topic Recommendation

The basic idea to recommend preferable topics for enjoyable conversation is discussed with detailed issues for previous methods.

2.1.1 Issues of Previous Methods

The issues of previous methods are as follows:

• Even if the users are grouped by using the user preferences, the system may not be able to recommend preferable topics for users.

• On the contrary, even if the topics are clustered by using similarity between topics, the system may not prepare the topic clusters which contain preferable topics for many users.

As a result, in the user group, there may be some users who do not have enjoyable conversation about the offered topics.

For example, we assume that users have the following user preferences:

- User A prefers about general soccer and politics.
- User B prefers about Japanese music and general sports.
- User C prefers about Japanese professional soccerleague and general music.
- User D prefers about general baseball and international politics.
- User E prefers about Japanese professional baseball and Japanese music.

We assume that user group (A, B, C) and (D, E) using user preference grouping method. For user A, B and C, they seem to enjoy their conversation with topics about Japanese professional soccer league, since the users have common preference of Japanese professional soccer league. For user D and E, they seem to enjoy their conversation with topics about Japanese professional baseball.

However, in case there are some topics about Japanother country relation and Japanese music, and no topics about Japanese

professional soccer league and Japanese professional baseball, the system cannot recommend appropriate topics along with their common preferences, so that they cannot have enjoyable conversation with the topics which the system collected in advance.

In the case of clustering topics previously, the similar issue may happen. We assume the following topics:

- Topic 1: A professional golfer showed his new hair style that is called "asymmetry."
- Topic 2: A soccer player with cool fashion went back to Japan.
- Topic 3: The Italian brand collaborates with an famous designer.
- Topic 4: The president of Italian premium brand said "The favorite shirt which was worn by the President in Japan is not absolutely my company product."
- Topic 5: Soccer manager participates in Ministry of Education, Culture, Sports, Science and Technology to improve sport status.

We assume that the topic cluster (1, 2, 3), which relates to fashion, and the topic cluster (4, 5), which relates to politics, are constructed using topic clustering method. In this case, if the topics about fashion are recommended to users who prefer fashion, and the topics about politics are recommended to users who prefer politics, their conversations seem to be stimulated. But, in case there are users who prefer sports or Italian culture, and no users who prefer fashion or politics, the system cannot recommend topics along with their common preferences, so that they cannot have enjoyable conversation with the topics.

Those issues come from handling of topics and user preferences separately. Since the topics and user preferences heavily rely on one another, the topics and user preferences should be handled together. Thus, for the example based on the user profile grouping, the user group should be modified as (A, D) and (B, C, E), since user A and D prefer politics, and user B, C and E prefer Japanese music. For the example based on the topic clustering, the topic clusters should be modified as (1, 2, 5) and (3, 4), since the users prefer sports or Italian culture.

2.1.2 Proposed Method

To handle the topic and the user preference together, The similarity is measured by using the topic and the user preference. For example, we assume that the user preference and the topic are expressed by form of word frequency vector. Then, the similarity can be measured by using $p_i \cdot t_j$, where p_i and t_j are the word frequency vectors for a preference of user i and topic j. Then, a topic preference vector w_j, based on the similarities, is expressed by the following form:

$$w_1 = (p_1 \cdot t_1, p_2 \cdot t_1, p_3 \cdot t_1, \cdots, p_N \cdot t_1),$$
$$w_2 = (p_1 \cdot t_2, p_2 \cdot t_2, p_3 \cdot t_2, \cdots, p_N \cdot t_2),$$

$$w_j = (p_1 \cdot t_j, p_2 \cdot t_j, p_3 \cdot t_j, \cdots, p_N \cdot t_j),$$

where N indicates the number of users.

If the topic preference vectors are clustered, based on the similarities, each cluster should contain appropriate topics which some users prefer, since those users have the large similarities to topics in the cluster. Then, those users corresponding the cluster are extracted using the degree of similarities, and the topics in the corresponding cluster are recommended to the extracted users.

2.2 Blueprint for Topic Recommendation System

Figure 1 shows a blueprint for the topic recommendation system. We assume that the system is used in real space, such as community centers and waiting rooms at hospitals. First, a user who come to the place is identified by robot, and the robot shows an appropriate topic recommendation area, where the topic recommendation area is linked to the topic cluster from which the user is extracted, Second, each user visits the area, and the appropriate topics are recommended by a robot to create opportunities of user conversation by talking to users, where we assume that the user is familiar with the robot rather than a personal computer. The robot throws the topics and comments which encourages user conversation about the topic, and sometimes asks user opinions. This communication stimulates users, and users at the same area can have conversations each other finally.

For example, we assume that there are an area recommendation robot and three topic recommendation areas, A, B and C. In

addition, there is a robot for the topic recommendation in the each area, and each topic recommendation area is linked to the topic cluster, respectively. When a user comes in front of the area recommendation robot, the robot indicates the topicrecommendation area A to the user, since the user is extracted from the corresponding topic cluster. After the area indication, the user moves to the topic recommended area A, and the topic recommendation robot in the area offers topics and comments, such as “The Japan team advanced to the finals at the Asia Cup 2011” and “I (the robot) think Japan team should be winner. Do the best. How does Taro think?.” Then, Taro speaks something like “Though Korea team also plays soccer well, I believed Japan team advanced to the final.”, and a neighbor person may respond to the utterance, such as “Yes. Current Japan team cannot lose. If there is a game near here in future, will you favor me with your company?.” Thus, the system gives chances to start conversation by gathering users with the same sort of topics.

2.3 Topic Recommendation System Configuration

Figure 2 shows the system configuration for the proposed topic recommendation system. To recommend topics, the system consists of two databases and three 3 construction step and one recommendation step.

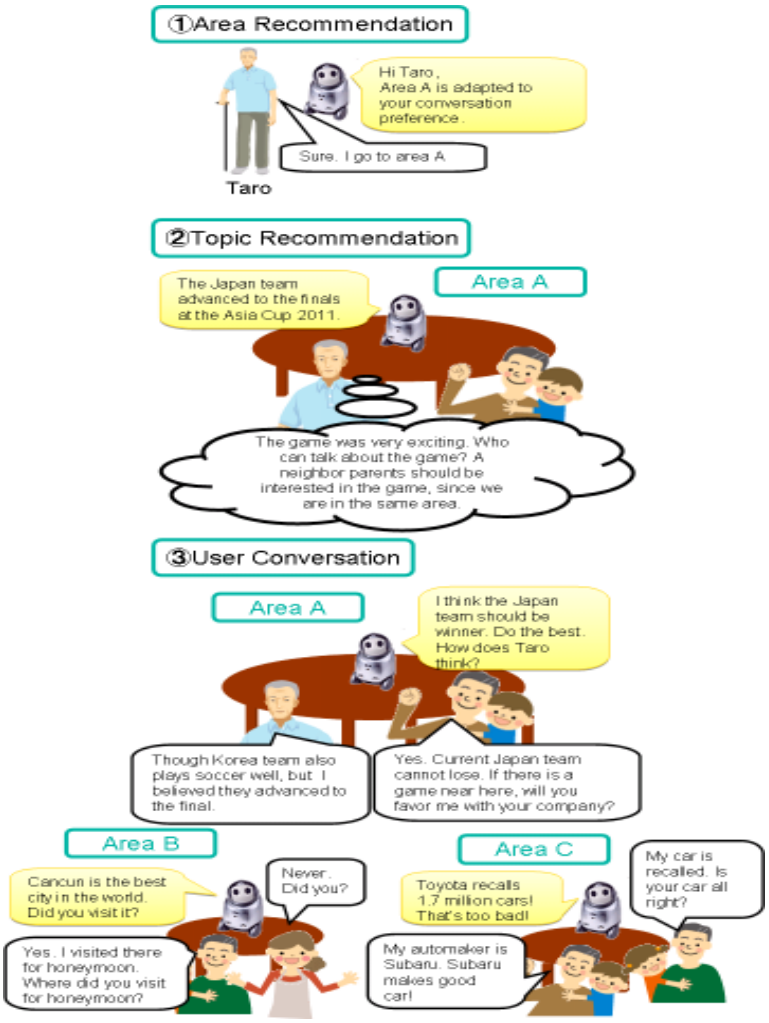

Figure 1. The blueprint of topic recommendation

(a) User Preference Database and Topic Database: The User preference database (DB) includes user preferences which are expressed in the form of word frequency vector, and the word frequency vector is called the user preference vector. The user preference vector is constructed from the user web browsing history. The topic DB contains topic IDs, its titles and contents. The contents are expressed in the form of word frequency vector, and the word frequency vector is called the topic vector. The topic DB may also contains the corresponding comments which the robot utilizes for talking to the users.

(b) Topic Preference Vector Construction: The topic preference vector is constructed bymeasuring similarities between the topic

vectors and the user preference vectors. The element of the topic preference vector is the user, and the value of the topic preference vector is the similarity. Using cosine similarity etc., the similarity between the topic vector and the user preference vector w_j is calculated.

(c) Topic Cluster Construction: The topic cluster is constructed by the topic preference vectors, based on the similarities, using the k-means algorithm, etc[5]. The number of clusters and the number of topics in the cluster depend on the size of available space and the number of users.

(d) User Extraction: The users are extracted from the topic cluster. Figure 3 shows an example to extract users from the topic clusters. Since the topic Figure 3: An example of appropriate user group construction for the corresponding topic clusters. preference vectors are clustered, based on the similarities, some users in the cluster indicates large similarities to the same sort of topics. Those users are extracted as a corresponding user group to the topic cluster. To extract those users, the topic preference vectors are summed in the topic cluster using $W_j = 1/n \sum_j w_j$, where n means the number of topics in the cluster. The users with the high similarities should be grouped for the corresponding cluster. The system extracts users in descending order of similarities for each cluster, and each user belongs to the corresponding topic cluster (In some cases, some users may be assigned to a number of topic clusters). This extraction is continued until all users belong to the topic cluster somewhere, and appropriate user groups are constructed for the topic clusters.

(e) Topic Recommendation: The appropriate topics should be recommended to the corresponding user groups. When a user visits the topic recommendation system, the system advises which user group is appropriate for the user. The user joins the user group, and the appropriate topics which corresponds to the user group are recommended by the robot.

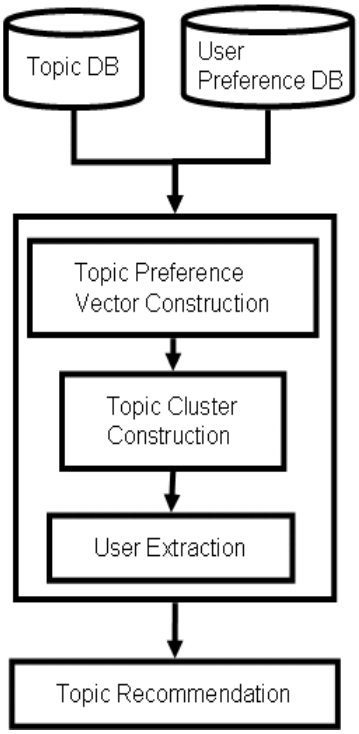

Figure 2. The system configuration

Figure 3. An example of appropriate user group construction for the corresponding topic clusters

3. Experiment and Evaluation Method

The proposed topic recommendation method was simply evaluated by offering topics and carrying out questionnaire survey to 10 users. The setup and the evaluation method for the simple experiment are explained in this section.

3.1 Experimental Participants and User Preference Vector

As a simple experiment before a large scale experiment, 10 employees in our company as users attended the experiment.

Since the users also attend a topic recommendation experiment in virtual space[4] in addition to this experiment, the topic recommendation experiment in virtual space is utilized to construct the user preference vector. The topic recommendation system in virtual space intends to provide chances to socialize with other users gradually by using topic exchanges. The users evaluate topics with "interested" or "not interested," and the system for the experiment in virtual space recommends other users who prefer similar topics to the user.

To construct the user preference vectors for this experiment, the noun words in the topics which are marked with "interested" are extracted and counted for each user.

3.2 Experimental Period

The experiment was conducted from May/25/2011 to Jun/15/2011. In the first nine days, topics are offered to users by the system for the experiment in virtual space to construct the user preference vector for each user. In the following 13 days, topics which are selected by the proposed system and topics for the comparisons are distributed to each user.

3.3 Conversation Partner

In the experiment, people did not meet together in real space actually. In stead, we assume that conversation is conducted by two persons: one is the user, another is the assumptive conversation partner. The following conversation partner is assumed:

• Unknown person (Partner0): Although the user exchanges greetings with the partner, the user knows little about the partner, and has little talk with the partner. For example, a person on the different floor in the same apartment.

• Known person, but may not know about the partner in details (Partner1): the user knows about some of the partner name, character, job, history etc., and has a little talk with the partner. But, the user does not know about the partner in details. For example, co-worker at the neighbor department.

• Well-known person (Partner2): the user knows about the partner in details (character, job, history, etc.). For example, childhood friends, familiar friends.

• Unknown person (Partner3): Although the user exchanges greetings with the partner, the user knows little about the partner, but the user knows that the person is also interested in the same sort of topics as the user. For example, a person with a dog in the same apartment (The user can guess the person likes dogs).

We carried out the questionnaire survey on the assumption of the conversation partner to the users. The three types, Partner0, Partner1, and Partner2, as the conversation partner are assumed to evaluate the topic recommendation methods, and the user evaluated topics on the assumption of conversations with each partner. Partner3 is assumed as the conversation partner to validate effectiveness of user group construction.

3.4 Offered Topics

In this experiment, a topic is defined as a text such as news article, weblog, etc., and is offered to users. The topics are collected from news site and famous weblogs in Japan, such as Yahoo! Japan News. The topics are offered as topic lists to users, and each topic list includes 50 topics. To validate the proposed topic recommendation method, the following types of topic lists are prepared:

• Randomly-selected topic list (TopicList0): The list includes topics which are randomly selected.

• Categorized topic list (TopicList1): The seven topic lists along with Yahoo! Japan News category is prepared[1], and each user picks one category from the categories at the user's discretion. The picked topic list is used as categorized topic list.

• Best topic list (TopicList2): The list includes topics which are deemed to be appropriate for the user by the proposed method.

• Worst topic list (TopicList3): The list includes topics which are deemed to be inappropriate for the user by the proposed method.

[1]There are seven categories, "Domestic News," "International News," "Business News," "Entertainment News," "Sports News," "Technology News," and "Local News" in Yahoo! Japan News.

o Use the topic aggressively.
Δ May use the topic, but not use aggressively.
× Not use the topic.
– Cannot evaluate since hard to understand the topic.

Table 1. Evaluation given by user for questionnaire, "Do you want to use the topic in conversation?"

o Want to know further details.
Δ Want to know, but if an opportunity offers.
× Be not interested at all.
– Cannot evaluate since hard to understand the topic.

Table 2. Evaluation given by user for questionnaire, "Do you want to know the topic in detail?"

The user evaluated these four topic lists including 50 topics, respectively, with the assumptive conversation partners to validate the proposed method.

3.5 Questionnaire and Evaluation Items

The user was asked the following questionnaires for the evaluations:

• Do you want to use the topic in conversation?

• Do you want to know the topic in details?

The evaluations to these questionnaires are given to each topic. The evaluation criteria are summarized in Table 1 and Table 2.

With the evaluations to the topics, the following simple items were evaluated:

• effectiveness of user group construction: Away from the proposed method, it is not clear whether the user group primarily activates conversation, even if the users who prefer the same sort of topics are grouped.

• completeness of user preference extraction: Although the user preference vector is constructed from web browsing history, that does not guarantee the complete extraction. Since the incompleteness of the user preference extraction may affect results for the proposed topic recommendation method, the incompleteness of the user preference extraction should be investigated.

• appropriateness of topic selection: The topics are clustered, based on the similarities between topics and the user preferences, by the proposed topic recommendation method. The topics should be investigated whether they are appropriate for the users.

	Partner0	Partner3
TopicList1	0.161	0.179

Table 3. The average ratio of the number of topics with o

TopicList0	0.096
TopicList1	0.108
TopicList2	0.174
TopicList3	0.255

Table 4. The average ratio of topics which users want to know in details

4. Results

In this section, effectiveness of user group construction, the completeness of user preference extraction and the appropriateness of topic selection are described.

4.1 Effectiveness of User Group Construction

Table 3 shows the average ratio of the number of topics with o for TopicList1. The ratio for each user is defined as ratio = $N_o / (50 - N_-)$, where N_o and N_- are the numbers of topics which are given o and –. Since the average ratio for Partner3 is higher than that for Partner0, the users in the same user group can maybe help to have more enjoyable conversation, if the user knows the conversation partner prefers the same sort of topics.

4.2 Completeness of User Preference Extraction

Table 4 shows the average ratio of topics which users want to know in details. Although TopicList2 and TopicList3 includes topics that was deemed to be appropriate and inappropriate using the similarity of word frequency vectors between the user preference and the topic by the system, TopicList3 has higher ratio, compared to TopicList2 and other ratios.

We confirmed what kind of noun words are used in TopicList2 and TopicList3, and compared themto noun words (user preference word list) which are used to construct user preference vectors. Table 5 shows the differences. The difference is defined as difference = $1 - N_B / N_W$, where N_B is the number of noun words which appear in both the TopicList and the user preference word list, and N_W is the number of noun words which appear in the user preference word list. More noun words in TopicList3, compared to TopicList2, did not appear in the user preference word list. This means that the user preference vector using web browsing history for nine days may not be enough to represent the whole user preference for topic recommendation.

	TopicList2	TopicList3
Difference	0.873	0.922

Table 5. Difference between noun words which appeared in TopicList and user preference word list

	Partner0	Partner1	Partner2
TopicList0	0.091	0.178	0.201
TopicList1	0.161	0.198	0.236
TopicList2	0.101	0.201	0.269

Table 6. Average ratio of the number of topics with o

4.3 Appropriateness of Topic Selection

Table 6 shows the average ratio of the number of topics with o. TopicList2, which contains topics selected by the proposed method, is appropriate for Partner1 and Partner2, but TopicList1 is more appropriate than TopicList2 for Partner0. Although the proposed method works for Partner1 and Partner2, the method does not work for Partner0. The following reasons are maybe considered:

• Since TopicList0 contains randomly selected topics, the user does not have knowledge about most of topics. As a result, it is difficult to use those topics in conversation.

• The user is scared to have no conversation, but also hesitate to talk along with his/her user preference, since the user does not know about character, preference, etc. for Partner0. Then, TopicList1 covers a wide range of topics from strongly preferring to just known for the user. TopicList1 contains many topics to be able to broaden conversation, so that the user prefers TopicList1 rather than TopicList2 for Partner0.

• It would be easy for the user to have conversation along with the partner preference, if the user knows some knowledge about the conversation partner. The user probably believes the conversation partner asks the user, if it is hard for the partner to follow

the topic. Then, the user can utilize topics in TopicList2 for Partner1 and Partner2.

5. Conclusion

To support senior citizen with the social problems, we proposed a method for prompting conversation through topic recommendation.

This method utilizes the similarities between topics and user preferences. Using the similarities, the topics are clustered, and the corresponding users are extracted from each topic cluster. Thus, this method can provide the appropriate topics to the corresponding users, and reduce the users who do not follow conversation due to uninterested topics as much as possible.

We conducted the simple experiment to validate effectiveness of the user group construction, completeness of the user preference extraction and appropriateness of topic selection. As a result, we confirms the following tendencies:

• the users in the same user group can help to have more enjoyable conversation, if the user knows the conversation partner prefers the same sort of topics,

• the user preference vector using web browsing history for nine days was not enough to represent the whole user preference for topic recommendation,

• the topics offered by the proposed topic recommendation method are appropriate for the known person, although the user preference extraction was not enough, but the topics offered along with a category are appropriate for the unknown person.

This time the simple experiment was conducted and some items are evaluated. We investigate effectiveness of this proposed topic recommendationmethod with a large scale experiment in near future.

6. Acknowledgment

This work has been supported by Japanese Ministry of Internal Affairs and Communications.

References

[1] Fujimoto, E., *et al*. (2004). *An* analysis of topic changes in free conversation using conceptual relations, *In*: The 18th Annual Conference of the Japanese Society for Artificial Intelligence.

[2] Sano, M., et al. (2009). A Robotic Introducer Agent Based on Adaptive Embodied Entrainment Control, p.368-376, *In*: Proc. HCI (2).

[3] Chen, J., et al. (2009). Make new friends, but keep the old: recommending people on social networking sites, p.201–210, *In*: CHI'09: Proc. of the 27th International Conference on Human factors in Computing Systems.

[4] Mizuguchi, H., et al. (2010). A Method for Building Friendship by a Networked Robot: Multi Stage Strategy of Recommendation, 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems.

[5] Jain, A. K., et al. (2010). Data Clustering: 50 years beyond K-means, *Pattern Recognition Letters* 31 (8) 651– 666.