# Improvement of recommendation list effectiveness using familiarity

J. Wu, T. Takayama, N. Sato, and Y. Murata

***Abstract***—This paper proposes to improve the effectiveness of recommendation items list by taking into account *familiarity* among customers. Recently, many shops are attracted to push type information provision that recommends their items to a customer. Collaborative filtering is one of the representative techniques for such purpose. However, it has mainly discussed recommendation precision, and satisfaction for an active user who is a customer who receives a recommendation has been not always sufficient. In general, it has possibility for a customer to be interested by themself in the item which a familiar person likes. Explicit example includes the case that a customer would like to try an item since their familiar person likes it. Implicit example includes the case that a customer has similar taste to their familiar friend and these two persons are attracted to similar item unconsciously.

In the present paper, we propose to introduce familiarity among customers into the conventional collaborative filtering in order to improve the effectiveness of recommendation list. More specifically, we propose the following two recommendation methods:

method 1): to use only the data of customers who are familiar with the active user, and

method 2): to provide each customer's data with adequate weight based on the familiarity to the active user.

We have conducted an evaluation experiment using our pilot system by relative comparison among the following three recommendations based on i) the method 1, ii) the method 2, and iii) general collaborative filtering. Its results have revealed that our proposition would be more effective than the general collaborative filtering. In the present paper, we also describe our consideration based on the *ANOVA* (ANalysis Of VAriance) to the result data in our experiment in order to enlarge the effectiveness.

***Keywords***—Collaborative filtering, effectiveness, familiarity, item recommendation.

## I. Introduction

RECENTRY, researchers have been aggressively studying recommendation techniques [1]-[4]. Especially, researches on improving the effectiveness of recommendation items list become more active. More specifically, approaches considering discovery ratio [5] or serendipity [6] are representative. In these approaches, customers' purchase history and/or their taste data are used as basis, however, the familiarity among customers is not fully considered. In general, it has possibility for an active user to be interested by themself in the item which their familiar customer likes. In the present paper, we try to improve the effectiveness of recommendation list by taking into account the familiarity among customers. We have conducted our evaluation experiment by utilizing television program as a product item like the reference [6]. Its results have revealed that our method would be effective.

The remainder of the present paper is organized as follows. Section II summarizes some related works as preliminaries of our discussion. Section III describes our proposed method that takes into account familiarity among customers. Section IV verifies the effectiveness of our proposition by evaluation experiment and compares it from the general collaborative filtering. Section V analyzes the result in the evaluation experiment, and describes some considerations in order to enlarge the effectiveness of our proposition. Finally, in Section VI, we presents conclusions and topics for future research directions.

## II. Related Work

### A. Collaborative Filtering

The basic idea of collaborative filtering is to recommend an item which similar user group in taste to an active user like, to the active user themself. Without taking into account the contents of the item, the collaborative filtering identifies some close customers only from the evaluation value which each user provides with each item. After that, it predicts an evaluation value of an item which the active user has not utilized yet.

We define some variables as follows.

- $U(=\{u_1, u_2, \ldots, u_a, \ldots, u_n\})$: $n$ customers who include the active user $u_a$.
- $I(=\{i_1, i_2, \ldots, i_m\})$: a set of $m$ items.
- $s_{ij}$: an evaluation value by a user $u_i$ to an item $i_j$.

Based on the above-mentioned definitions, similarity $\rho_{ai}$ of an active user $u_a$ to another user $u_i$ is calculated by the *Pearson correlation formula* (1) which concerns the item evaluated in common.

J. Wu is with Graduate School of Software and Information Science, Iwate Prefectural University, Iwate 020-0193, Japan (e-mail: g231j201@s.iwate-pu.ac.jp).

T. Takayama is with Graduate School of Software and Information Science, Iwate Prefectural University, Iwate 020-0193, Japan (corresponding author to provide phone: +81-19-694-2614; fax: +81-19-694-2614; e-mail: takayama@iwate-pu.ac.jp).

N. Sato is with Graduate School of Software and Information Science, Iwate Prefectural University, Iwate 020-0193, Japan (e-mail: nobu-s@iwate-pu.ac.jp)

Y. Murata is with Graduate School of Software and Information Science, Iwate Prefectural University, Iwate 020-0193, Japan (e-mail: y-murata@iwate-pu.ac.jp)

$$\rho_{ai} = \frac{\sum_{k \in y_{ai}} (s_{ak} - \overline{s_a'})(s_{ik} - \overline{s_i'})}{\sqrt{\sum_{k \in y_{ai}} (s_{ak} - \overline{s_a'})^2} \sqrt{\sum_{k \in y_{ai}} (s_{ik} - \overline{s_i'})^2}} \quad (1)$$

Here,

- $y_{ai}$: a set of items which are commonly evaluated by $u_a$ and $u_i$.
- $\overline{s_{i'}}$ and $\overline{s_{a'}}$ : the average evaluation value by the user $u_i$, and $u_a$ for commonly evaluated items, respectively.

Using the formula (1), $s_{aj}$, which is an evaluation value of the user $u_a$ to the item $i_j$, is predicted by formula (2), as a weighted average of the evaluation value to item $i_j$ by each user. In here, $x_j$ implies a set of users who have already evaluated item $j$, and $\overline{s_a}$ implies average evaluation value of items which have already evaluated by the user $u_a$.

$$s_{aj} = \overline{s_a} + \frac{\sum_{i \in x_j} \rho_{ai}(s_{ij} - \overline{s_i'})}{\sum_{i \in x_j} |\rho_{ai}|} \quad (2)$$

### *B. Researches Improving Effectiveness of Recommendation List*

Hijikata *et al.* [5] propose a collaborative filtering algorithm taking into account discovery ratio. It introduces a binary measure: 'already-known' or 'unknown' item for a user, and calculates similarities among some users and ones among some items. Based on them, it predicts some items which an active user is 'unknown' but likes, and improves the effectiveness of recommendation list.

Murakami *et al.* [6] propose a method for improving the serendipity of recommendation. It introduces 'custom model' in addition to the taste model, and treats an item which may be accessed in custom as low serendipity one. More specifically, it improves the effectiveness of a recommendation list by removing 'low serendipity items based on the custom model' from the recommendation list made of the taste model.

However, in the previous researches, we cannot always say that familiarity among customers is fully considered.

### *C. Social Media and Recommendation*

Recently, researches on trying to effectively utilize a knowledge obtained from a social media have been activated.

Budak *et al.* [7] propose trend detection method for online social network. More specifically, they classify a trend into two categories by 'coordinated' and 'uncoordinated', in other words, clustered and distributed users. They also propose a sampling technique for structural trend detection. Their experiments performed on a Twitter data set show that even with a sampling rate of 0.005, the average precision to detect trend is 0.93 for coordinated trends and 1.00 for uncoordinated trends.

Freyne *et al.* [8] propose to recommend a relevant feed, based on the action that an active user had taken on a social network in the past, such as view, friending, browsing, and interacting with other users. More specifically, they evaluate regularity of each action. If a feed that usually taken a positive action happens, such as a feed by a good friend, they recommend it.

Gürsel *et al.* [9] develop recommendation method for posted photos on a photo sharing social website (*flickr.com*). They classify each photo into suitable category based on tags, provide reliability per each category, and determine recommended item based on metadata and comments

However, these two researches [8]-[9] recommend a posted content itself. On the other hand, our research is different from them in the following point. That is, our goal is to recommend an item which talked on a social network, not a contents itself fed on the social network.

Machanavajjhala *et al.* quantitatively analyze a trade-off relation between privacy violation and recommendation accuracy, in case of personalizing a recommendation based on a social media [10]. The more we hide evidence data of a recommendation, the more each person's privacy is protected. However, it has possibility to become a recommendation lost an active user's interest. On the other hand, it is not desirable to provide evidence data of a recommendation without any restriction, for example, whose data the evidence is. Therefore, we intentionally show that a recommendation is based on the data by an active user's familiar person in order to lead them to try its recommended item.

## III. Proposed Method

Recently, in social media including 'Twitter' or 'Facebook', people are attracted to represent their own opinion and react to them. In this research, we assume that we can obtain both each user's taste and familiarity data among the users, from their social media. In the present paper, we represent familiarity '$F_{ai}$' between an active user $u_a$ and the other user $u_i$ in five levels from '5: good' to 'bad: 1'. Level 3 indicates 'neutral'.

In here, the above-mentioned assumption: "we can obtain familiarity data among the users from their social media" is not unrealistic. We carry out our discussion with summarizing some such researches.

Yang *et al.* [11] model 'link prediction/ friendship prediction' and 'interest targeting/ service recommendation' that have close relation in *SNS*, using a framework 'FIP (Friendship-Interest Propagation) model'. It leads to obtain familiarity data among the users from their social media. However, they have only achieved to a binary link prediction: exist or not. It can be only applied until a limitation that it recommends an item which a person who predicted link to an active user likes. They have not achieved to a level that each customer is ranked quantitatively based on analysis. We discuss in the viewpoint that how we can take advantage of its obtained data in the case where we represent each familiarity in multiple stage level based on their idea.

Actually, Gilbert *et al.* [12] propose a model to predict strength of tie between two users from *Facebook*. According to their experimental results, they have achieved 85.0% precision.

They conclude that their model can be applied into 'social media design', 'privacy controls', 'message routing', 'friend introduction', and 'information prioritization'. As an additional application area, we investigate 'item recommendation'.

One important thing to be investigated is to determine the attitude for privacy preservation when we obtain familiarity data between users from social media. Wu *et al.* [13] study to find strength of official and/or private tie among employees based on behavioral analysis on a local *SNS* in a single company. Recently, like some companies such as *Best Buy, Deloitte, Microsoft, and IBM*, *SNS* bounded in a company is often used in order to activate communication and/or collaboration inside the company. Although these *SNS* inside company is hard to protect anonymousness, on the other hand, it is known that it has not a small advantage in strategic team composition and effective placement of human resource. Therefore, we hold the deep discussion of anonymousness to another paper. We investigate how much effectiveness we can obtain when we pick up a familiarity among customers from a social media and take advantage of it to item recommendation.

We come back to the discussion to recommendation method in the present paper. We propose the following two types of recommendation methods: method 1 and 2.

### A. Method 1: Only Using Familiar Customers' Data

This method sets a threshold of the familiarity value. It utilizes only familiar users' data with larger or equal to familiarity value for an active user than the threshold, in the use of formula (1) and (2). For example, if we adopt familiarity 3 as its threshold, data of neutral or more familiar person to an active user is utilized for collaborative filtering.

### B. Method 2: Weight Provision Based on Familiarity

It provides each user with their importance based on familiarity for an active user. More specifically, it first obtains a normalized familiarity $F_{ai}'$ from 0.0 to 1.0, based on the familiarity $F_{ai}$.

$$F_{ai}' = \frac{F_{ai} - 1}{4} \tag{3}$$

Second, it multiplies $F_{ai}'$ to right hand side of the formula (1), and utilizes it as a weight $\rho_{ai}'$. Third, it replaces $\rho_{ai}$ in the formula (2), which predicts the active user's taste, to $\rho_{ai}'$.

$$\rho'_{ai} = F_{ai}' \times \rho_{ai}$$
$$= \frac{F_{ai}' \times \sum_{k \in y_{ai}} (s_{ak} - \overline{s_a}')(s_{ik} - \overline{s_i}')}{\sqrt{\sum_{k \in y_{ai}} (s_{ak} - \overline{s_a}')^2} \sqrt{\sum_{k \in y_{ai}} (s_{ik} - \overline{s_i}')^2}} \tag{4}$$

## IV. Evaluation Experiments

We use a television (TV) program as a product item like the reference [6], and create a recommendation list. Hereafter, we often abbreviate 'TV program item' to 'program item' or 'item'.

The subjects are total 31 persons who are under-graduate or graduate student belonging to the laboratory of IT. In general, satisfying diversity is desirable in subject selection. However, since the proposed method assumes the existence of familiar persons, we use the above-mentioned subjects. We set enlargement of subjects to future work.

### A. The First Survey: Determination of Items Used in Our Experiment

#### A.1 Experiment Method

In the first survey, we determine the items used in our experiment. More specifically, we investigate a TV program item that the subjects often watch. Since there exists lots of TV program items even in a single day, we select some items which we collect the data of 'watch or not' or 'evaluation value of favorite'. We ask three questions in the first survey.

Fig. 1 shows the questionnaire form of Q1. Its question is "Which genre of TV program do you watch? Please check the appropriate option (Multiple answers allowed)". Twenty options are provided based on *EPG* (Electronic Program Guide). This genre classification is flat and relatively fine-grained. In order to make each subject easy to answer, we adopt this classification in here. Last option: 'others (specifically: )' is provided for the case where a subject feels that all options do not fit to them. Answers to Q1 clarify exhaustiveness of genre in selected subjects.

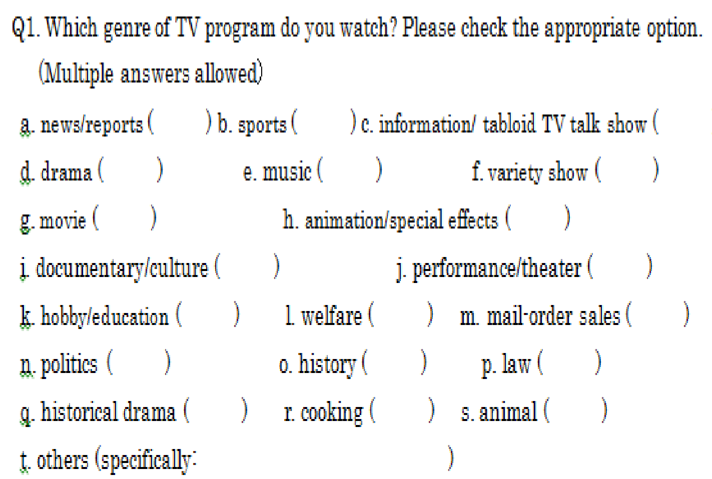
Q1. Which genre of TV program do you watch? Please check the appropriate option.
(Multiple answers allowed)
a. news/reports ( ) b. sports ( ) c. information/ tabloid TV talk show ( )
d. drama ( ) e. music ( ) f. variety show ( )
g. movie ( ) h. animation/special effects ( )
i. documentary/culture ( ) j. performance/theater ( )
k. hobby/education ( ) l. welfare ( ) m. mail-order sales ( )
n. politics ( ) o. history ( ) p. law ( )
q. historical drama ( ) r. cooking ( ) s. animal ( )
t. others (specifically: )

Fig. 1 contents of Q1

Fig. 2 shows the questionnaire form of Q2. Its question is "Please write the TV programs which you watch and/or record almost every time, up to five". It is main question in the first survey, and investigates which program item each subject usually watches. Its response table includes 'name of TV program', 'TV station name', 'day of the week', and 'time'.

Q2. Please write the TV programs which you watch and/or record almost every time, up to five.

| name of TV program | TV station name | day of the week | time |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Fig. 2 contents of Q2

Fig. 3 shows the questionnaire form of Q3. It asks demographic information (graduate/undergraduate, grade, sex, age, and name (optional)) of each subject.

Q3. Please tell us about you.

graduate/ undergraduate: (circle one) grade:

sex (answer 'male' or 'female'): age name (optional)

Fig. 3 contents of Q3

*A.2 Experiment Result*

Table 1 shows the results of Q1. Genres excepting 'performance/theater', 'welfare', and 'law' are exhausted.

Table 1 results of Q1

| genre code | genre name | # of person |
|---|---|---|
| a | news/reports | 26 |
| b | sports | 11 |
| c | information/ tabloid TV talk show | 11 |
| d | drama | 8 |
| e | music | 6 |
| f | variety show | 17 |
| g | movie | 14 |
| h | animation/special effects | 15 |
| i | documentary/culture | 4 |
| g | performance/theater | 0 |
| k | hobby/education | 7 |
| l | welfare | 0 |
| m | mail-order sales | 1 |
| n | politics | 4 |
| o | history | 1 |
| p | law | 0 |
| q | historical drama | 5 |
| r | cooking | 3 |
| s | animal | 1 |
| t | other | 0 |
| | total | 134 |

These results are referenced when we determine a set of television programs used in our evaluation experiment.

Table 2 shows the results of Q2. Although the total number of TV program items we have obtained is 143, they are 38 items after we exclude overlaps among them. Five items which include four animations and one variety show have already been finished to be broadcasted. Hereafter, we adopt thirty-three items subtracted the five from the thirty-eight ones, for our experiments. We note that some items belong to multiple genres such as 'news/reports' and 'information/ tabloid TV talk show'.

We also note that we map the original twenty genres classification to twelve ones in Table 2. This new classification is based on the reference [14]. Although we have investigated the exhaustiveness of genre in Q1, it has been not easy to select TV program items as exactly fitting to it. Therefore, as the efforts we can make, we adopt more rough genre classification than the one in Q1. Its new classification is hierarchical, and these genres belong to the most rough granularities. Since all of the old fine-grained twenty genres are included in the hierarchical tree of the new classification, we can map the old genre into a new one.

Table 2 the numbers of programs per genre

| genre name | # of program |
|---|---|
| news/reports | 2 |
| sports | 2 |
| information/ tabloid TV talk show | 1 |
| drama | 1 |
| variety show | 6 |
| movie | 2 |
| animation/special effects | 15 |
| hobby/education | 1 |
| news/reports and information/ tabloid TV talk show | 5 |
| drama, and music | 1 |
| variety, and performance/theater | 1 |
| music | 1 |
| total | 38 |

Table 3 shows the results of Q3. Although we conduct our experiments with these subjects, increase of subjects is one of the future works for us.

Table 3 subject's demographic information

| grade | # of person | male | female |
|---|---|---|---|
| G 2 | 2 | 2 | 0 |
| G 1 | 3 | 2 | 1 |
| UG 4 | 9 | 6 | 3 |
| UG 3 | 8 | 5 | 3 |
| UG 2 | 9 | 9 | 0 |
| total | 31 | 24 | 7 |

G=graduate, UG=undergraduate

### *B. The Second Survey: Data Collection of 'Each Evaluation Value for Item' and 'Each Familiarity between Users'*

#### *B.1 Experiment Method*

As we have described in the section III, we suppose that we can obtain the information of 'each user's taste' and 'familiarity among users' from social media. In the present paper, in order to obtain these information more easily, we conduct the second survey.

In Q1 of the second survey, we use a questionnaire form shown in Fig. 4, and ask each subject to write each evaluation value for thirty-three items determined in the first survey. We use six levels of evaluation value:

- '5: like it very much',
- '4: like a little',
- '3: know it, however, neither like nor dislike it,
- '2: do not like it very much',
- '1: do not like it', and
- '*: do not know it'.

Q1. Please circle the appropriate evaluation value for each TV program.

<evaluation value>

5. like it very much
4. like a little
3. know it, however, neither like or dislike
2. do not like it very much
1. do not like it
*. do not know it

| genre | name of TV program | the station name | day of the week | evaluation value |
|---|---|---|---|---|
| news/ reports | NHK news (morning) | NHK | Monday - Friday | 5 4 3 2 1 * |
| | news watch 9 | NHK | Monday - Friday | 5 4 3 2 1 * |
| sports | F1 grand prize | 'Menkoi' TV | Sunday | 5 4 3 2 1 * |
| | F1 preliminary round of grand prize | 'Menkoi' TV | Saturday | 5 4 3 2 1 * |
| information/ tabloid TV talk show | 'Miyame' shop | 'Nihon' TV | Monday - Friday | 5 4 3 2 1 * |
| … | … | … | … | … |

Fig. 4 contents of Q1 in the second survey

In Q2, we ask a familiarity $F_{ai}$ of user $u_i$ for a user $u_a$ by two measures: 'frequency to meet' and 'density of relationship' (Fig. 5).

First, 'frequency to meet' has five levels including

- '5: almost every day',
- '4: three or four days per a week',
- '3: one or two days per a week',
- '2: sometimes/occasionally', and
- '1: no acquaintance/ almost no talk even if two persons meet each other'.

It does not enforce the existence of conversation.

Second, 'density of relationship' has six levels:

- '5: very good relationship/ have lots of talk',
- '4: good relationship/ have a talk moderately',
- '3: normal relationship/ have a little talk',
- '2: not good relationship/ do not talk too much',
- '1: bad relationship/ almost no talk, and
- '*: no acquaintance'.

Q2. Please circle the most appropriate answer code.

| < frequency to meet (no conversion is OK) > | < density of relationship > |
|---|---|
| 5. almost everyday | 5. very good relationship/ have lots of talk |
| 4. 3 or 4 days per 1 week | 4. good relationship/ have a talk moderately |
| 3. 1 or 2 days per 1 week | 3. normal relationship/ have a little talk |
| 2. sometimes/occasionally | 2. not good relationship/ do not talk too much |
| 1. no acquaintance/ almost no talk even if 2 persons meet each other | 1. bad relationship/ almost no talk |
| | *. no acquaintance |

| grade | name (omitted titles) | frequency to meet | density of relationship |
|---|---|---|---|
| G 2 | K. K | 5 4 3 2 1 | 5 4 3 2 1 * |
| G 2 | H. S | 5 4 3 2 1 | 5 4 3 2 1 * |
| G 1 | A. C | 5 4 3 2 1 | 5 4 3 2 1 * |
| G 1 | S. S | 5 4 3 2 1 | 5 4 3 2 1 * |
| G 1 | J. W | 5 4 3 2 1 | 5 4 3 2 1 * |
| UG 4 | S. I | 5 4 3 2 1 | 5 4 3 2 1 * |
| UG 4 | Y. I | 5 4 3 2 1 | 5 4 3 2 1 * |
| … | … | … | … |

G=graduate, UG=undergraduate

Fig. 5 contents of Q2

In this experiment, we calculate $F_{ai}$ by arithmetic mean of these two data: 'frequency to meet' and 'density of relationship':

$$F_{ai} = \frac{(frequency\_to\_meet)+(density\_of\_relationship)}{2} \tag{5}$$

As an exception in the formula (5), if 'density of relationship' is '*', we set $F_{ai}$ to zero. Although we write omitted titles at the second column from the left on the table in Fig. 5, we actually need to write each full name, not the omitted title. This survey has possibility to violate a subject's privacy. We explain this survey's purpose and how to use the obtained data, and ask to reply in the possible range.

### B.2 Experiment Result

Table 4 shows the evaluation value distribution of item in Q1. Since '*' is most ratio, we could say that there would exist enough room in order to recommend unknown items to a subject. The second most is '3: know it, however, neither like nor dislike it', and all evaluation value is exhaustive.

Table 4 evaluation value distribution of item in Q1

| evaluation value | # of program | ratio (%) |
|---|---|---|
| 5 | 93 | 9.1 |
| 4 | 106 | 10.4 |
| 3 | 255 | 24.9 |
| 2 | 52 | 5.1 |
| 1 | 160 | 15.6 |
| * | 357 | 34.9 |
| total | 1023 | 100.0 |

Table 5-7 shows the results of Q2. All evaluation values are exhaustive in Table 5-7, and it is desirable for our experiments. We note that there is no case 'familiarity' = 0.5 in the Table 7, because $F_{ai} = 0$ when 'density of relationship' = *.

Table 5 frequency distribution to meet

| frequency to meet | # of person | ratio (%) |
|---|---|---|
| 5 | 116 | 12.4 |
| 4 | 231 | 24.8 |
| 3 | 198 | 21.3 |
| 2 | 208 | 22.4 |
| 1 | 177 | 19.1 |
| total | 930 | 100.0 |

Table 6 density distribution of relationship

| density of relationship | # of person | ratio (%) |
|---|---|---|
| 5 | 149 | 16.0 |
| 4 | 201 | 21.6 |
| 3 | 205 | 22.1 |
| 2 | 134 | 14.4 |
| 1 | 166 | 17.8 |
| * | 75 | 8.1 |
| total | 930 | 100.0 |

Table 7 familiarity distribution to other user

| familiarity | # of person | ratio (%) |
|---|---|---|
| 5.0 | 41 | 4.4 |
| 4.5 | 86 | 9.2 |
| 4.0 | 142 | 15.3 |
| 3.5 | 110 | 11.8 |
| 3.0 | 238 | 25.6 |
| 2.5 | 62 | 6.7 |
| 2.0 | 73 | 7.8 |
| 1.5 | 55 | 5.9 |
| 1.0 | 48 | 5.2 |
| 0.0 | 75 | 8.1 |
| total | 930 | 100.0 |

### C. The Third Survey: Evaluation of Proposition

The purpose of this third survey is to evaluate our proposition.

### C.1 Experiment Method

Fig. 6 shows the questionnaire form. We ask each subject to evaluate each top five items of the following three recommendation lists based on:

- general collaborative filtering,
- method 1, and
- method 2.

In Fig. 6, each value of second column from the left is inserted depending upon the data obtained until the second survey. They vary depending upon a subject.

Q. How much would you like to watch each TV program?
Select one of the following.
5. would like to watch it very much.
4. would like to watch it.
3. neither would like to watch nor wouldn't.
2. would not like to watch it very much.
1. would not like to watch it.

1. Recommendation List by not taking into account TV programs which your familiar persons watch

| recommendation rank order | program ID | this time evaluation value | next time evaluation value |
|---|---|---|---|
| 1 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 2 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 3 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 4 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 5 | | 5 4 3 2 1 | 5 4 3 2 1 |

2. Recommendation List 1 by taking into account TV programs which your familiar persons watch

| recommendation rank order | program ID | this time evaluation value | next time evaluation value |
|---|---|---|---|
| 1 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 2 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 3 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 4 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 5 | | 5 4 3 2 1 | 5 4 3 2 1 |

3. Recommendation List 2 by taking into account TV programs which your familiar persons watch

| recommendation rank order | program ID | this time evaluation value | next time evaluation value |
|---|---|---|---|
| 1 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 2 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 3 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 4 | | 5 4 3 2 1 | 5 4 3 2 1 |
| 5 | | 5 4 3 2 1 | 5 4 3 2 1 |

Fig. 6 contents of 3rd questionnaire survey

Fig. 7 shows a recommendation list page in our pilot system. On the screen, we call a recommendation list based on the general collaborative filtering as ‘Recommendation List by not taking into account television programs which your familiar persons watch’. As the same way, we call a recommendation list based on the method 1 as ‘Recommendation List 1 by taking into account television programs which your familiar persons watch’. Therefore, we call a recommendation list based on the method 2 as ‘Recommendation List 2 by taking into account television programs which your familiar persons watch’. We note that we shuffle the provision order of the three recommendation list fairly, because we avoid the order effect.

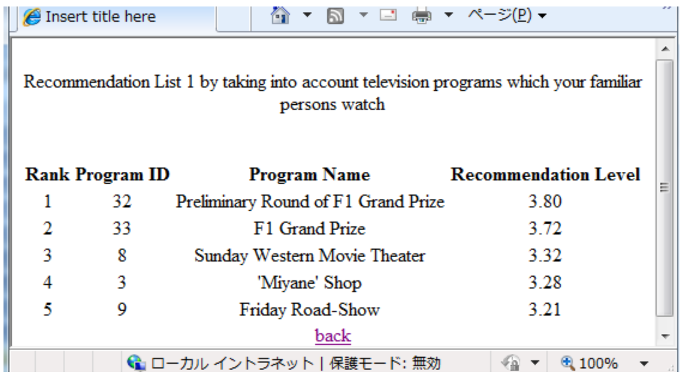


Fig. 7 recommendation list page in our pilot system

We show each subject following the three information:

a) program information in the form of paper media based on *EPG*,
b) a preview video concerning after a set of *CM* (Commercial Message). It is usually shown just before the start of *CMs*, and
c) a preview video of the next time. It is usually shown in the end of a television program.

After that, we ask to write “how much would you like to watch each television program in the five levels?” ‘This time evaluation value’ should be determined from the above-mentioned a), b), and ‘next time one’ from c). After that, we treat the arithmetic average of these two evaluation values as a level that how much the subjects would like to watch them. We note that when we find the item already evaluated in other recommendation list, we copy and reuse the value.

We use the video of the television program of three parts:

1) the preview of the next time,
2) the preview concerning after a set of *CM*, and
3) a beginning of the program,

in this priority order (Fig. 8). The upper limit of the video length is thirty seconds in total. We set the above priority because a television program often does not have ‘the preview video concerning after a set of *CM*’ and/or ’the preview video of the next time’. Actually, Table 8 shows the results in investigating whether a television program used in our experiment contains the above-mentioned 1) and/or 2).

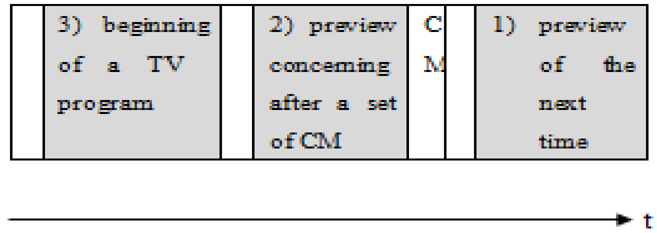


Fig. 8 classification of the TV program video along time axis

Table 8 existence of ‘the preview concerning after a set of *CM*’ and ‘preview of the next time’

| | preview concerning after a set of CM | preview of the next time |
|---|---|---|
| Yes | 19 | 20 |
| No | 14 | 13 |

We evaluate the effectiveness of each recommendation list by *DCG* (Discounted Cumulative Gain) [15]. Although *DCG* is originally a measure used for evaluation of search engine, recently, it is also used for evaluation of recommendation list.

$$DCG(i) = \begin{cases} G(1) & (if \;\; i=1) \\ DCG(i-1) + \dfrac{G(i)}{\log_2 i} & (otherwise) \end{cases} \tag{6}$$

In here, *G(i)* is relevance of item ranked order ‘*i*’. The purpose of the second formula is to let the weight light, when the ranking order becomes bad. It leads to the following: the more high relevant item is placed in the good ranking order, the more the value of *DCG* becomes big. Analysis using *DCG* often overstates each relevance. In the present paper, we adopt the reference [16]’s overstating method which translates relevance ‘2’ and ‘1’ into ‘0’.

*C.2 Experiment Result*

Fig. 9 shows the results in analysis by *DCG*. It compares average *DCG* per method. We can observe that the recommendation list based on the method 1 and method 2 are more effective than the general collaborative filtering. There is no significant difference between the method 1 and method 2. We describe further considerations in the next section.

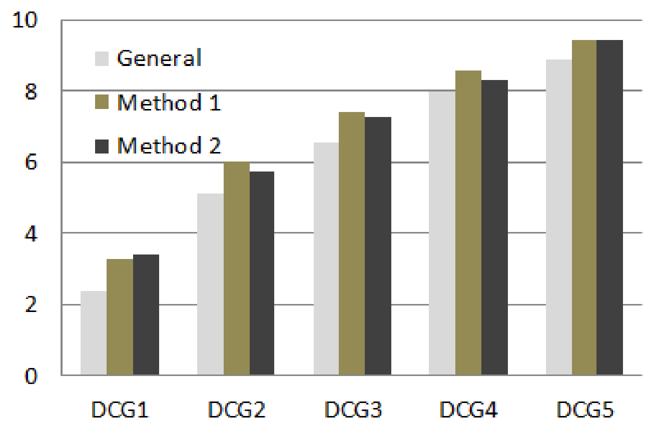


Fig. 9 the evaluation result of the three recommendation list by *DCG*

## V. Considerations

### A. Effectiveness Analysis of Proposed Method

Table 9 shows the results in statistical significance test by *ANOVA* (ANalysis Of VAriance) [19] to Fig. 9. We can observe significant difference in only *DCG*(1), except for *DCG*(2)-(5). For actual use, we had better obtain significant difference in *DCG*(2)-(5), too. Therefore, in the next subsection, we investigate the method of enlarge its difference.

Table 9 results of *ANOVA* for Fig. 9

*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 4.83 | 1.01 | 1.10 | 0.49 | 0.52 |

### B. Towards Enlargement of Effectiveness

In this subsection, we proceed with our discussion in the following policy. That is, after we bring each idea, we test its effectiveness by relative comparison of *DCG* in newly arising a set of three recommendation lists. In other words, before we change our pilot system, we simulate 'change of ranking order' and 'effectiveness' brought by each new idea. We reuse the data obtained from the three surveys in the section IV.

#### B.1 Against the Inclination of Genre

Table 2 in the subsection IV.A shows that the numbers of item belonging to 'animation/ special effects' is relatively big. Hereafter we abbreviate its genre name to '*anim*'. Although we actually decreased the numbers to eleven since four items had already been finished to broadcast, it is still relatively high ratio compared to the other genres. When such an *anim* item arises in a recommendation list, it has possibility to have been degraded evaluation value in a subject who does not like *anim*. Therefore, we investigate the evaluation values, in the case where we exclude the influence of *anim*. More specifically, we analyze in the following three methods.

(1): To replace the evaluation value of *anim* item into 3.2, which is the average value among total.
(2): To delete *anim* item in each recommendation list, and lift the following items to the better ranking order direction. In this approach, in some cases, vacancy arises on the lower ranking order. We calculate average value of *DCG* based on only occupied data excepting vacancy.
(3): Partial use of *anim* data. Using the popularity ranking of *anim*, we adopt about half well-known items from the total *anim* ones selected for the experiments. More specifically in the present paper, we adopt the evaluation values of top five *anim* items in the popularity ranking. After that, we analyze in the following two policies.
  (3-1): We replace the evaluation values of worse order item than ranking order five into average evaluation value: 3.2.
  (3-2): We delete worse order item than ranking order five, and lift the following items to the better ranking order direction.

First, we concretely begin our discussion about above-mentioned (1). Fig. 10 shows the results in *DCG* of this case. Compared from Fig.9, differences between 'general' and 'method 1', and ones between 'general' and 'method 2' become small.

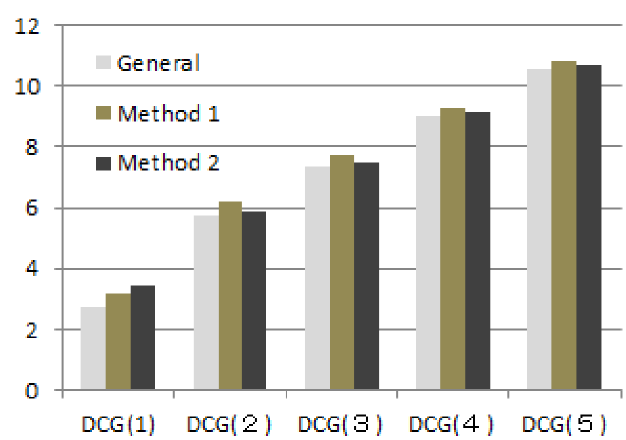


Fig. 10 *DCG* in case of replacing evaluation value for *anim* item into 3.2

Table 10 shows the results in statistical significance test by *ANOVA* to Fig. 10. As overall tendency, *F-value* becomes smaller than Table 9. Even in *DCG(1)*, we can not observe significant difference.

Table 10 results of *ANOVA* for Fig. 9

*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 3.37 | 0.63 | 0.88 | 0.35 | 0.35 |

Second, we concretely discuss (2): 'to delete *anim* item in each recommendation list, and lift the following items to the better ranking order direction'. Fig. 11 shows the results in *DCG* of this case. We can observe that the method 2 is still superior than the general collaborative filtering. However, the method 1 is inferior to the general collaborative filtering in *DCG*(4), (5).

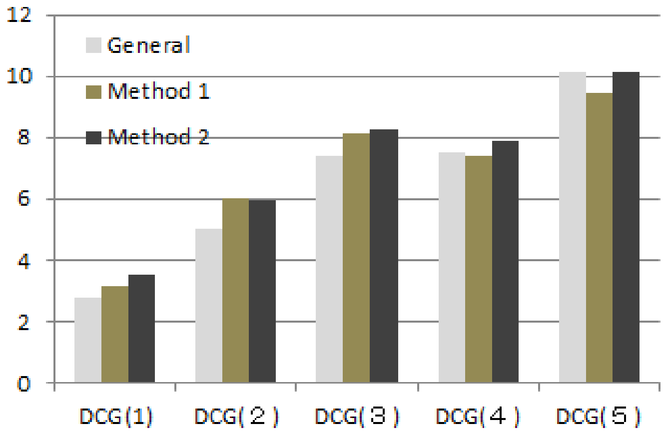


Fig. 11 *DCG* in case of deleting *anim* items and lifting the worse ranked items

Table 11 shows the results in statistical significance test to Fig. 11. Also in this case, we can not observe significant difference between the general collaborative filtering and our proposition: the method 1, 2.

Table 11 results of *ANOVA* for Fig. 11
*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 1.38 | 1.11 | 0.03 | 0.19 | 0.16 |

Third, we concretely discuss (3): 'partial use of *anim* data'. We adopt the reference [18] as the popularity ranking of *anim*. It's target period is from Oct. to Dec., 2011 and fits to our experiments. The total numbers of voting is 8547, and the numbers of voting person is 3173. It is up to twenty-seven ranking order, and contains five items of the eleven ones which we had used in our experiments.

When we adopt the policy (3-1): 'we replace the evaluation values of worse order item than ranking order five into 3.2', Fig. 12 shows the results in *DCG*. In this case, we can observe that both the method 1 and 2 is clearly better than the general collaborative filtering in all of *DCG*(1)-(5).

Table 12 shows the results in statistical significance test to Fig. 12. Although significant difference exists only in *DCG*(1) as the same as in the Table 9, the *F-values* in *DCG*(2)-(5) have become better than in the Table 9.

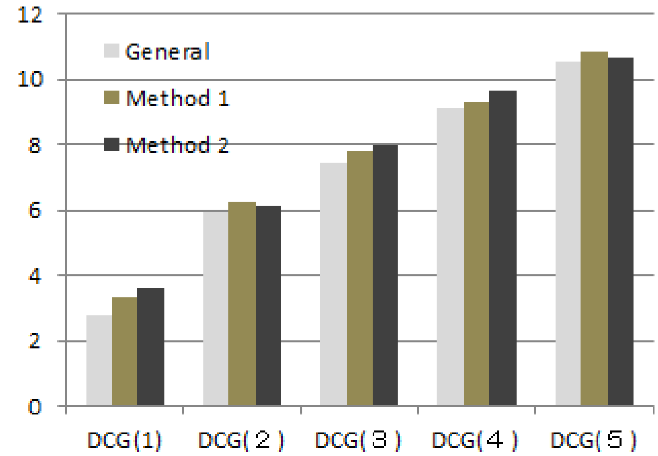


Fig. 12 the case we adopt the evaluation value of popularity ranking top five *anim* item, and replace one to each *anim* under popularity ranking five into average evaluation value: 3.2

Table 12 results of *ANOVA* for Fig. 12
*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 3.90 | 1.84 | 3.02 | 1.86 | 1.65 |

When we adopt the policy (3-2): 'we delete worse order item than ranking order five, and lift the following items to the better ranking order direction, Fig. 13 shows the results in *DCG*. In this case, we can not say that differences between 'general' and 'method 1', and ones between 'general' and 'method 2' become large compared from Fig. 9.

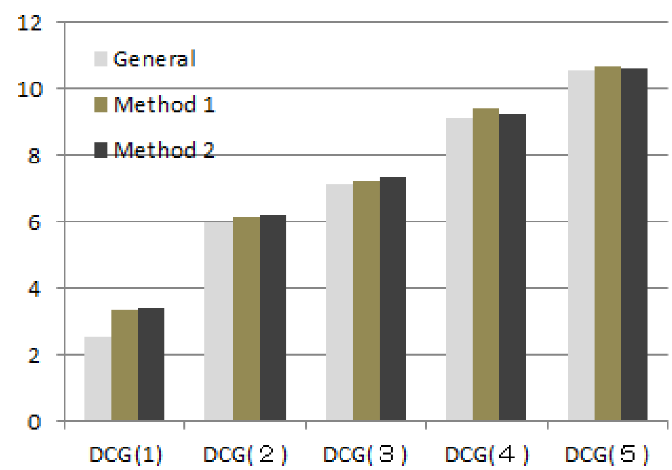


Fig. 13 the case we adopt the evaluation value of popularity ranking top five *anim* item, delete each *anim* item under popularity ranking five, and lift the remained contents

Table 13 shows the results in statistical significance test to Fig. 13. We can not observe any significant difference through *DCG*(1)-(5).

Table 13 results of *ANOVA* for Fig. 13

*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 0.45 | 0.33 | 0.13 | 0.02 | 0.01 |

As a consideration, concerning the inclination of *anim* genre, the (3-1) policy is most effective. If we adopt about half well-known items from the total *anim* ones selected for the experiments, it has possibility to improve the effectiveness of our proposed methods.

### *B.2 Modification of Overstating Method in DCG Analysis*

As we had described in the subsection IV.C.1, we adopted the overstatement method in the reference [16]. We investigate the case where we do not translate each relevance from '2' and '1' into '0' in *DCG* analysis.

Fig. 14 shows the results in *DCG*. In this case, we can observe that both the method 1 and 2 is clearly better than the general collaborative filtering in almost all of *DCG*(1)-(5).

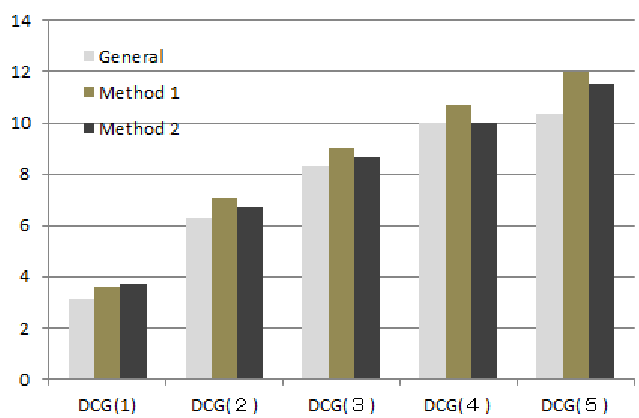


Fig. 14 results in *DCG* of the case where we do not translate each relevance from '2' and '1' into '0'

Table 14 shows the results in statistical significance test to Fig. 14. We can observe significant difference in *DCG*(1)-(3), (5), except for *DCG*(4). It means that we have already obtained statistically significant difference that our proposed methods are more effective than the general collaborative filtering, if we adopt the *DCG* analysis without overstatement. In order to obtain the effectiveness not depending upon an overstating method, further modifications of proposed methods are desirable.

Table 14 results of *ANOVA* without overstatement in *DCG* analysis

*Boundary value of F ratio at the 5% level of significance = 3.39*

| DCG | DCG(1) | DCG(2) | DCG(3) | DCG(4) | DCG(5) |
|---|---|---|---|---|---|
| F-value | 5.01 | 3.44 | 5.75 | 2.06 | 13.17 |

## VI. Conclusion and Future Research Directions

In the present paper, we have proposed to improve the effectiveness of recommendation list, by taking into account the familiarity among customers. We have conducted our evaluation experiment using our pilot system, and obtained the results that our method would be more effective than the general collaborative filtering. According to an *ANOVA*, the effectiveness of our proposition would have statistically significant difference from the general collaborative filtering.

We are planning some future researches: i) to enlarge subject in the evaluation experiment, ii) further modification of the proposed methods in order to enlarge the effectiveness, iii) relative comparison of our method from the existing approaches considering discovery ratio or serendipity, and iv) to investigate concerning introduction of fuzzy logic.

## References

[1] M. Corniel, F. Gil, J. Molero, J. Ferrer, A. M. Borges, R. Gil, L. Contreras, "Studies orientation and recommendation system (SORS): use case model and requirements," NAUN International Journal of Systems Applications, Engineering & Development, Issue 3, Volume 5, 2011, pp.387-395.

[2] T. Y. Lai, G. H. Hwang, "Application of half-life theory and fuzzy theory to a selection and recommendation system for web advertisement delivery in consideration of the time effect," WSEAS Transactions on Information Science and Applications, Issue 1, Volume 7, 2010, pp. 70-80.

[3] Taowei Wang, Yibo Ren, "Research on personalized recommendation based on web usage mining using collaborative filtering technique," WSEAS Transactions on Information Science and Applications, Issue 1, Volume 6, 2009, pp. 62-72.

[4] A. Toma, R. Conatantinescu, F. Nastase, "Recommendation system based on the clustering of frequent sets," WSEAS Transactions on Information Science and Applications, Issue 5, Volume 6, 2009, pp. 715-724.

[5] Y. Hijikata, T. Shimizu, and S. Nishida, "Discovery-oriented collaborative filtering for improving user satisfaction," Proceedings of the 14th ACM International Conference on Intelligent User Interfaces(ACM IUI 2009), 2009, pp.67-76.

[6] T. Murakami, K. Mori, and R. Orihara, "Metrics for evaluating the serendipity of recommendation list," New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science, Springer, Vol.4914, 2008, pp.40-46.

[7] C. Budak, D. Agrawal, A. E. Abbadi, "Structural trend analysis for online social networks," Proceedings of the 37th International Conference on VLDB (VLDB2011), 2011, pp. 646-656.

[8] J. Freyne, S. Berkovsky, E. Daly, and W. Geyer, "Social networking feeds: recommending items of interest," Proceedings of the 4th International Conference on ACM Recommender Systems (RecSys2010), 2010, pp. 277-280.

[9] A. Gürsel, S. Sen, "Producing timely recommendations from social networks through targeted search," Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS09), 2009, pp. 805-812.

[10] A. Machanavajjhala, A. Korolova, A. D. Sarma, "Personalized social recommendations - accurate or private?," Proceedings of the 37th International Conference on VLDB (VLDB2011), 2011, pp. 440-450.

[11] S. H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, H. Zha, "Like like alike - joint friendship and interest propagation in social networks," Proceedings of the 20th International Conference on WWW (WWW2011), 2011, pp. 537-546.

[12] E. Gilbert, K. Karahalios, "Predicting tie strength with social media," Proceedings of the 27th International Conference on ACM CHI (CHI2009), 2009, pp. 29-38.

[13] A. Wu, J. M. DiMicco, D. R. Millen, "Detecting professional versus personal closeness using an enterprise social network site," Proceedings of the 28th International Conference on ACM CHI (CHI2010), 2010, pp. 1955-1964.

[14] Association of Radio Industries and Businesses, *Standard Sequence Information for Digital Broadcasting System*, ARIB STD-B10 Ver. 4.6, 2008.
[15] K. J¨arelin,. and, J. Kek¨al¨ainen., "IR evaluation methods for retrieving highly relevant documents," Proc. 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2000, pp.41–48.
[16] K. Oku, S. Nakajima, J. Miyazaki, S. Uemura and H. Kato, "A ranking method based on users' contexts for information recommendation," Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication(ICUIMC 2008), 2008, pp.289-295.
[17] J. Levandoski, M. Ekstrand, M. Ludwig, A. Eldawy, M. Mokbel, J. Riedl, "RecBench, benchmarks for evaluating performance of recommender system architecture," Proceedings of the 37th International Conference on VLDB (VLDB2011), 2011, pp. 911-920.
[18] Animation popularity ranking by maniac animation fans, Available: http://flagmamire.lolipop.jp/rank/ranki.html
[19] A. E. Watkins, *Statistics in Action, Analysis of Variance and Anova CD Package : Understanding a World of Data*, John Wiley & Sons Inc, 2008.