Title	Collaborative Filtering versus Personal Log based Filtering: Experimental Comparison for Hotel Room Selection
Author(s)	Saga, Ryosuke; Tsuji, Hiroshi
Citation	
Issue Date	2005-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/3865
Rights	©2005 JAIST Press
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, IFSR 2005 : Proceedings of the First World Congress of the International Federation for Systems Research : The New Roles of Systems Sciences For a Knowledge-based Society : Nov. 14-17, 2075, Kobe, Japan, Symposium 2, Session 3 : Creation of Agent-Based Social Systems Sciences Formal Systems

Collaborative Filtering versus Personal Log based Filtering: Experimental Comparison for Hotel Room Selection

Ryosuke Saga[1] and Hiroshi Tsuji[2]
[1] Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka, Japan
saga @mis.cs.osakafu-u.ac.jp
[2]Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka, Japan
tsuji@ cs.osakafu-u.ac.jp

ABSTRACT

In order to support decision making in e-commerce domain, recommender system has attractive features such as collaborative filtering and personal log based filtering for products/services. As experimental study, this paper compares these filtering for hotel room selection. Differently from commodity items, products/services as hotel rooms have three features: many attributes, multiformity and high-frequency update. Noting that we cannot use explicit rating data assigned by users, this paper describes how to derive implicit rating from sales records. Numerical simulation shows how accuracy between two filtering exists, where our case data consist of 10,000 users, 400,000 personal log and 160,000 room plans.

Keywords: Intelligent Agent, Recommender System, Information Filtering, TPO-Goods, Collaborative Filtering.

1. INTRODUCTION

Recently, an intelligent agent works to help user selection in e-commerce domain. It often recommends the preferable goods/services [1]. Collaborative filtering is one of the representative techniques for the domain and is applied to recommend the several items such as CD and books [2][3].

At the same time, TPO (means Time, Place, and Occasion)-goods such as hotel rooms and airline tickets exist as another commodity items [4][5]. For TPO-goods, we have developed the technique called log based filtering [5]. While collaborative filtering has been proved useful for non-TPO-goods [6][7], its applicability to TPO-goods has not been known yet.

In order to evaluate the accuracy of collaborative filtering to TPO-goods with simulation, this paper compares it to that of log based filtering. In this simulation, we use actual hotel room data.

The reminder of this paper is structured as follows. Section 2 reviews the features of TPO-goods and discusses the issue including explicit rating and implicit rating to predict user's preference. Section 3 describes the log based filtering and collaborative filtering, where we also present F-measure algorithm by similarity calculation. Section 4 shows the simulations of log based filtering and collaborative filtering. Finally, section 5 concludes this paper.

2. ISUUES FOR RATING USER'S PREFERENCE ON TPO-GOODS

2. 1. Features of TPO-goods

Attributes in TPO-goods such as hotel and airline ticket are sensitive to external factors [4][5]. The factors represent season, location and *event* related goods although we may list several factors.

TPO-goods have the three features which depend on the external factors: The first feature is that the number of attribute is high. For example, if we reserve a hotel room, we should check not only rate but also distance from mass transit, room size, service and so on. The second feature is multiformity derived from several combinations of the attributes. The last feature is that the external factors force to update attributes of TPO-goods, which is the most remarkable point.

2. 2. Consideration of recommender system

In order to recommend goods/services, an information filtering system, called recommender system, should rate user's preference. Generally, there are two kinds of rating method: explicit rating and implicit rating [5]. The former is consciously rated by users in form of description as like semantic differential methods used on GroupLens [3]. The latter is not expressed by users but is recorded in database as log. Web visiting log and sales record are typical examples.

Because rates for TPO-goods are often time-variant, its recommender system should use the implicit rating. An explicit rating for goods at one TPO is not the same as for the same goods at different TPO. For an example, a rate for resort hotel at on season cannot be the same value at off season.

Therefore, we prefer implicit rating to explicit rating for recommender system for TPO-goods. Then let us discuss how to use the personal sales records which is the typical implicit data for recommender system.

3. RECOMMENDER SYSTEM FOR TPO-GOODS

There are two typical types in recommender system: content-based filtering and collaborative filtering where our log based filtering relates to content-based filtering.

3. 1. Personal Log based filtering

Personal log based filtering (or log based filtering) is developed as recommender system for TPO-goods [5]. This underling concept is borrowed from content-based filtering [7][8] and log based filtering makes recommendation based on user profile derived from transaction log such as personal sales records.

In order to create user's profile, sales records work statistics analyisis. Then the patterns resulted from the analysis is expressed as distribution. Let us call this distribution preference distribution and the value on it preference value. Then it is possible to predict the preference of items by refering to the preference distribution whether the attribute value is recorded or not. In this paper, the preference value of the attribute j on item x is defined as pj(x) whose range is from 0 to 1. At the case pj(x) =1, the value of attribute on x is regarded as most preferable.

Creating user profile, this recommender system maps the goods catalog into the preference distribution. The system makes recommendation from the preference distribution by three search patterns: high-angle search,

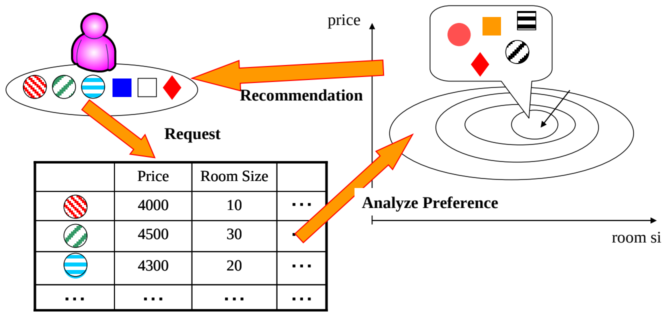


Figure 1 Personal Log based Filtering

low-angle search, and neighbor search (Figure 2). The first pattern, high-angle search, searches items from the most preferable area for user. The second pattern, low-angle search, makes recommendation by searching items from the selected goods to the preferable area and the third pattern, neighbor search, finds the items around the selected goods without preference distribution.

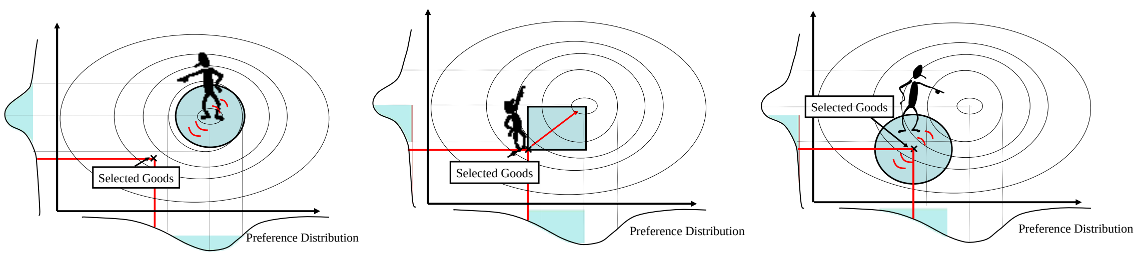
3. 2. Collaborative filtering

Collaborative filtering [2][3][6] exists as a concept against content-based filtering. The basic premise of collaborative filtering is that similar users might like similar things. Therefore, this filtering makes recommednation based on rating by other users.

The basic processes of collaborative fitltering consists of following steps:

1. To identify the similar users on their preference,
2. To recommend items which thet preferred.

Especially, the first step is important among above steps because of the basic premise. The method used on GroupLens is the representive example called correlation algorithm. In GroupLens, user rates the articles by 5 level, then the system identifies the similarity user based on Peason correlation [10] using the rating. However, correlation algorithm depends on



(a) High-Angle Search (b) Low-Angle Search (c) Neighbor Search

Figure 2 Recommendation Patterns

explicit rating and is not appropriate for TPO-goods.

Here, we show sales records into Venn diagrams. Venn diagrams shows correlation of two sets by the degree of intersection, where the set shows the sales records and the intersection shows items set which users bought in common. As is obvious, the intersection tends to expand according to buying the same items (see Figure 4), and we can regard the correlation as the similarity because the tendency is similar to that of similarity.

F-measure [11], which is used for the measurement of retrieval performance, has the same tendency of the correlation in Venn diagram. Therefore, we can measure the similarity of users by regarding sales records as document set. F-measure is calculated by the harmonic mean between recall and precision which are also the measurement. When we define the recall and precision as formula (1), F-measure is shown as formula (2)

$$R = \frac{|I_a \cap I_b|}{|I_a|}, P = \frac{|I_a \cap I_b|}{|I_i|} \qquad (1)$$

$$F = \frac{(1+\beta^2)RP}{\beta^2 R + P} \qquad (2)$$

where R and P shows recall and precision, $|I|$ shows the element number of set I, and I_a and I_b are the set of rating (that is, sales record) by user a and user b. β has more than zero, which shows importance degree of recall compared with precision. Incidentally, the recall for user a is regarded as the precision for user b, that is, the relation between R and P is symmetric relation, and the importance of recall is able to regard as equal to that of precision. Therefore, we set β to 1 and the formula (1) is transformed into the following simpler formula:

$$sim(a,b) = F - measure_{\beta=1} = \frac{2RP}{R+P} \qquad (3)$$

4. SIMULATION

4. 1. Simulation environment

The goal of simulation is to compare log based filtering with collaborative filtering. In the simulation, we use the actual data of business hotel provided by BestReserve Co.,Ltd [12] who serves internet based business hotel room reservation. We use the data which consists of about 10,000 users who have reserved more than 25 times between July/2000 and April/2004, their 400,000 sales records, and 160,000 room plans in stored database.

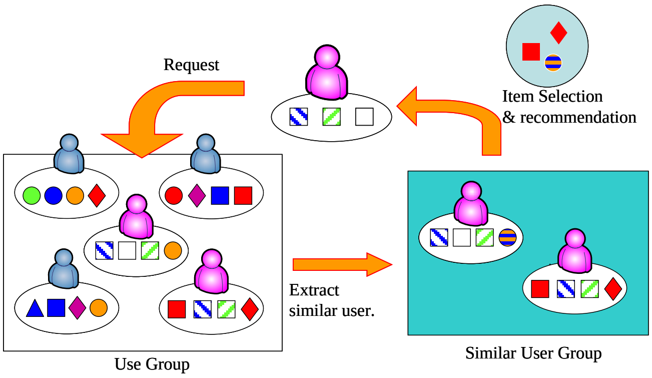


Figure 3 Collaborative filtering

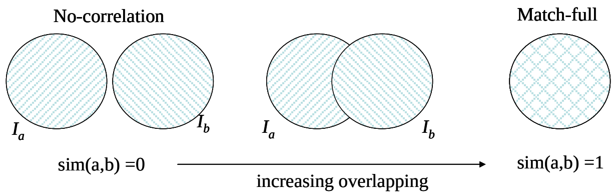


Figure 4 Relationships between Increasing Similarity and Expanding Intersection

As criteria, we use the following parameters because they are the most highlight when user makes reservation in the web page: price, room size, distance from mass transit, and breakfast service.

At first, using these data, we carry out the simulation of log based filtering where we use three patterns high-angle search, low-angle search, and neighbor search. Here, the selected goods for recommendation are generated by changing the sales records. Next, we carry out the simulation of collaborative filtering. Note that it must take account of the variation of attributes by TPO. In log based filtering, the agent retrievals items based on individual preference and recommends preferable goods although the attributes on items changes frequently. However, in collaborative filtering, the recommended items are not changed because collaborative filtering depends on the items which are bought and evaluated by other person in spite of changing the attributes.

Consequently, we assume three cases: on season, off-season, and the other season. And as corresponding to each case, we configure the three price patterns: the case of highest price, the case of lowest case, and the case of average price. Figure 5 shows the simulation flow of collaborative filtering.

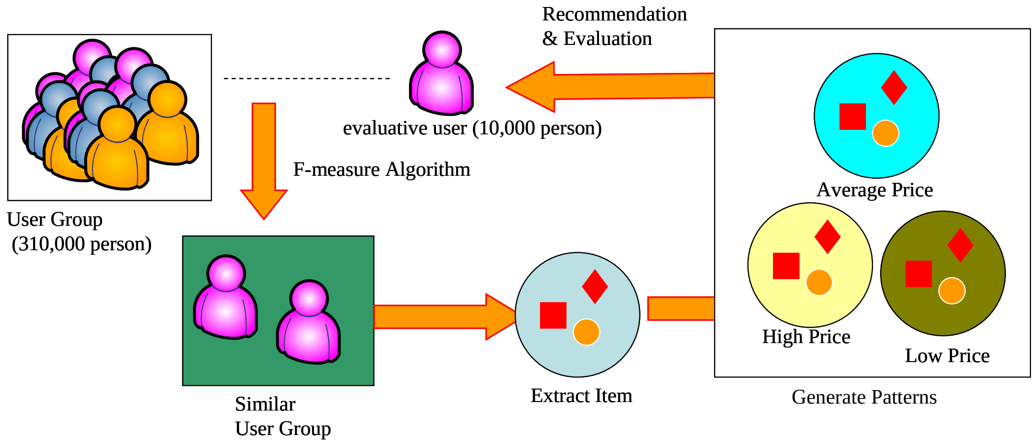


Figure 5 Simulation Flow of Collaborative filtering

Additionally, we use goods fitness value for criteria. Goods fitness is evaluated value based on the preference extracted sales records and is shown as formula (4)

$$(goods \quad fitness) = \sum_{j \in K} p_j(x) \qquad (4)$$

where K is the set of attributes. In this simulation, the maximal value is 4 because the number of attribute is 4.

4. 2. Simulation Result

The simulation results are shown in Figure 6 and Figure 7. In the figures, the horizontal axis expresses the number of sales records, and the vertical axis expresses the average of goods fitness.

As the result, the accuracies of high-angle search and low-angle search is 3~3.5 as shown in Figure 6. The accuracy on the case of average price also is almost equal to them as shown in Figure 7. In the other hands, the accuracies of the other cases is lower than above cases. We can guess the reason as follows:

- In neighbor search, the preference distributions is not used as we mentioned at above section. The affection leads to low fitness value.
- In the two cases of collaborative filtering, not only price but also other services such as breakfast service are changed, and each of them counteracts each other. As the result, the recommeded goods become to undesirable for users (Figure 8).

5. CONCLUSION

In this paper, we have carried out the comparison simulation between personal log based filtering and collaborative filtering targeting hotel room selection and have checked the accuracy.

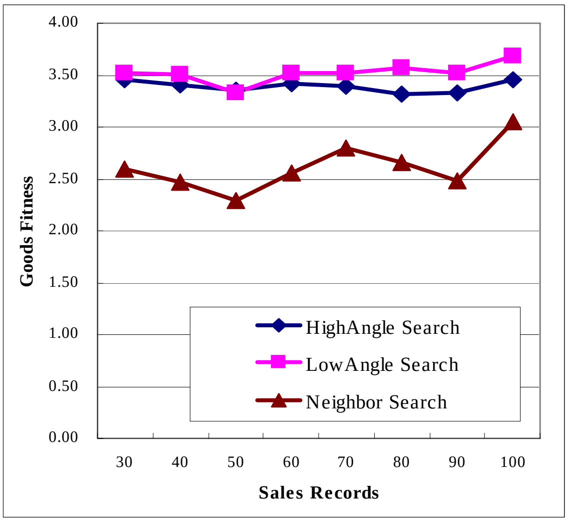


Figure 6 Simulation Result of Personal Log based Filtering

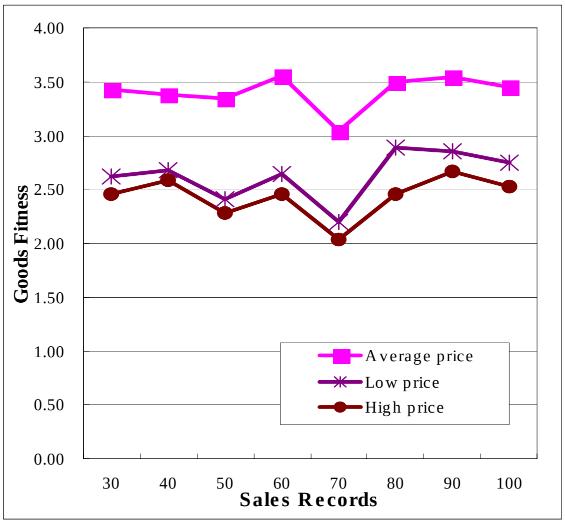


Figure 7 Simulation Result of Collaborative Filtering

Differently from commodity items, TPO-goods as hotel rooms have three features: many attributes, multiformity and high-frequency update and we have shown that we could not use explicit rating data assigned by users for recommender system.

The numerical simulation has shown that the accuracy of log based filtering except neighbor search kept high performance, and the accuracy of collaborative filtering was lower than them and changed by TPO. Therefore, we have found that personal log based filtering is more appropriate for the hotel room selection than collaborative filtering in this case study.

ACKNOWLEDGEMENT

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Exploratory Research, 17650073, 2004-2005.

Additionally, authors would like to express sincere thanks to Mr. Jun Onoda, CEO of BestReserve Ltd., who had given us chance to use transaction data for experimentation.

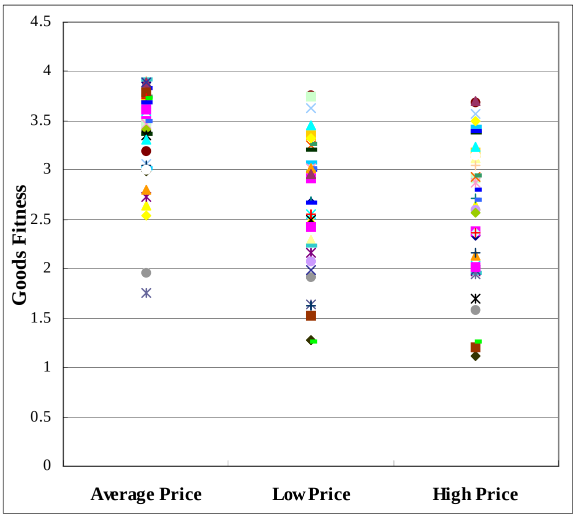


Figure 8 Scatter Graph of Recommended Goods Fitness

REFERENCES

[1] P. Resnick and H. Varian, “Recommender System”, *Communications of the ACM*, Vol.40, No.3, pp. 56-58, 1997.

[2] D. Goldberg, D. Nichols, D. Gupta, and C. Perkins, ”Using collaborative Filtering to weave an information Tapestry”, *Communications of the ACM*, 35(12), pp. 61-70. 1992.

[3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: An Open Architecture for Collaborative Filtering of Netnews”, *Proceedings of Conference on Computer Supported Cooperative Work (CSCW'94)*, Chapel Hill, NC, ACM Press,175-86, 1994.

[4] R. Saga and H. Tsuji, “Alternative Goods Recommender by Customer’s Context -Example for Business Hotel Room Reservation-“, *Proceeding of The Third International Conference on Active Media Technology*, pp.331-334, 2005.

[5] R. Saga, H. Tsuji, and J. Onoda, “Agent System for Notifying Hotel Room Reservation Alternatives”, *11th International Conference on Human-Computer Interaction, 2005* (in CD-ROM).

[6] J. S. Breese, D. Heckerman and C. Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering”, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp.43-52, 1998.

[7] M. Montaner, B. López and J.Ll de la Rosa, “A Taxonomy of Recommender Agents on the Internet”, *Artificial Intelligence Review 19*, pp. 285-330, 2003.

[8] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation", Communications of the ACM, vol. 40, pp. 66-72, 1997.

[9] B. Krulwich, “LIFESYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data”, *Artificial Intelligence Magagine 18(2)*, 37-45 (1997)

[10] C.J.V. Rijsbergen, Information retrieval, Second edition, Butterworths, 1979.

[11] G. Salton and M. McGill, Introduction to Modern Information Retrieval. New York, NY: McGraw-Hill Publishing Company,1983.

[12] BestReserve Co.,Ltd, http://www.bestrsv.com/