Improvement of Combination Information Filtering Method based on Reliabilities

Yusuke Ariyoshi
Human Media Research Laboratories, NEC Corporation,
8916-47, Takayama-cho, Ikoma, Nara, 630-0101, Japan

Abstract

In this paper, a hybrid filtering method, which combines the content-based filtering (CBF) method and the social information filtering (SIF) method, is proposed. The proposed method generates final recommendations that have higher precision by using the reliabilities of the recommendations generated by CBF and SIF. The hybrid method calculates the reliability of the intermediate recommendations based on the quantity of learned ratings and words. It then, merges the intermediate recommendations. This in turn changes the weights of CBF and SIF, which are based on their reliabilities, and generates the final recommendation. Through experiments, it was found that SIF contributed to strengthened learning in CBF. This works very well even when the number of users is small.

1. Introduction

There are currently two major types of information filtering methods. One is the CBF (Content-Based Filtering) method, which uses actual content features, and the other method is the SIF (Social Information Filtering) method (or so-called collaborative filtering), which uses other users' content ratings. CBF can filter information that has not been previously evaluated. SIF, on the other hand, can filter information that contains figures and tables that are hard for CBF to analyze.

Recently, combinations of CBF and SIF have been researched to develop a method that has the advantages of both CBF and SIF. These combination methods can be classified into two types: prediction combination and feature combination. The prediction combination method merges predictions generated by each filtering method [Ariyoshi and Ichiyama, 1997; Balabanovic and Shoham, 1997]. The feature combination method integrates content features and social features before filtering [Basu and Hirsh, 1997; Billsus and Pazzani, 1998].

Basically, the prediction combination method uses CBF to predict scores of unrated information and SIF to predict scores of rated information. However, this method did not adequately investigate how to combine CBF and SIF. For example, if there are few users who have rated specific information, the CBF prediction of the information is more reliable than the SIF prediction.

In this paper, a new combination method called the "hybrid method" is proposed. This introduces the reliability of the predictions of CBF and SIF into the prediction combination method. The hybrid method first calculates prediction reliability based on the quantity of learned ratings and words. Then, the weights of CBF and SIF are adjusted based on these reliabilities in order to merge the final predictions. The conventional prediction combination method strengthens the learning of word importance in CBF by using SIF predictions. In contrast, highly reliable predictions are only added to strengthen learning by using reliabilities in the hybrid method.

The next section describes the conventional filtering methods and their weaknesses. Section 3 discusses the hybrid method. Section 4 presents the results of experimental tests, and section 5 discusses the conclusions to this.

2. CBF and SIF

This section explains and compares CBF and SIF. In this paper, the ratings used are based on numeric scores given by users.

2.1 CBF

CBF filters information based on matching information content with user's interests. In many cases, information content and user's interests are represented by vectors that consist of the importance of terms.

In a typical CBF [Nakamura *et al.*, 1997], information is defined by document vectors that represent the presence/absence of terms. In the document vector, the value 1 indicates the presence of a specific term in a document, and 0 indicates absence. The representation of a user's interest, which is called a profile, is a vector that represents the value of terms to a user.

CBF consists of a score prediction and a learning interest (Fig.1).

Score prediction: Suppose $\boldsymbol{p}$ is a current profile vector and $\boldsymbol{x}$ is a document vector of information **x**, then the prediction value **S'x** of information **x** is defined as the inner product of $\boldsymbol{p}$ and $\boldsymbol{x}$.

Learning interest: Upon receiving the real score **Sx** at **x**, vector $\boldsymbol{p}$ is updated to $\boldsymbol{p'}$ that gives a partition of the segment **S'x Sx** with a rate of α:(1-α). Here, the parameter α is called the learning rate and determines how conservative the algorithm is.

2.2 SIF

People who have similar interests often recommend useful information to each other. Conceptually, SIF is an automation of the information sharing that is often done by word of mouth. In a typical SIF, a vector of a user's ratings on recommended information represents a user's interests. SIF consists of a user-similarity calculation and a score prediction (Fig.2).

User-similarity calculation: In this step, the similarity between users is measured by using the correlation coefficient of their past ratings. For example, the similarity **r**(u, v) between users **u** and **v** is computed as:

$$r(u,v)=\frac{\sum_{x:\text{Information rated by } u \text{ and } v}\left(S_{ux}-\overline{S_u}\right)\left(S_{vx}-\overline{S_v}\right)}{\sqrt{\sum_{x:\text{Information rated by } u \text{ and } v}\left(S_{ux}-\overline{S_u}\right)^2}\sqrt{\sum_{x:\text{Information rated by } u \text{ and } v}\left(S_{vx}-\overline{S_v}\right)^2}}.$$

In the above formula, **Sux** is a score that is rated by user **u** on information **x**. Here, $\overline{Su}$ is the average of user **u**'s ratings.

Score prediction: Predictions are generated by computing a weighted average of other users' ratings. The similarities between the current user and other users are used as weights. The weighted average uses the ratings of users who have similarities greater than a certain threshold [Shardanand and Maes, 1995]. For example, the predicted score **S'ux** for user **u** on information **x** is computed as:

$$S'_{ux}=\frac{\sum_{v:|r(u,v)|>\beta}|r(u,v)|S_{vx}}{\sum_{v:|r(u,v)|>\beta}|r(u,v)|}.$$

2.3 Comparison of CBF and SIF

As described above, CBF filters information based on the information content (terms), and SIF filters information based on scores given by other people. Because of these differences, the two methods have the following characteristics:

I) Importance of figures and tables: In CBF, information importance is determined by word terms. Thus, it is difficult to recommend information whose importance is in figures or tables. In principle, it is possible to analyze images. Though, to analyze importance of images is still difficult. Here, in contrast, SIF is likely to produce quality recommendations since the importance of the information has already been assessed by other users.

II) New information: CBF is able to filter information that has not been evaluated by other people. In contrast, SIF can only recommend information that has been evaluated by other users.

III) Unlearned terms: CBF learns a user's profile only through information that has been rated by the user himself/herself. Therefore, CBF accuracy declines if information contains many unlearned terms that have not previously appeared in rated information. The ability to filter new information is one strength of CBF. However, many important new terms (such as "Java" or "XML") appear on the Internet every day. Because CBF can not understand new terms means that it cannot filter these new topics well. This can be a weakness of CBF when it is used in information recommendation services.

2.4 Combination Method

Recently, combinations of CBF and SIF have been researched. There are two major types of combination methods: prediction combination and feature combination. The prediction combination method merges predictions generated by each filtering method. In this method [Ariyoshi and Ichiyama, 1997] has two combination stages. One merges prediction, and the other strengthens learning (Fig.3).

The prediction merging stage integrates the predictions of CBF and SIF. Basically, CBF is used for predicting scores of new information that has not been rated, and SIF is used for predicting scores for information that has been rated.

The strengthened learning stage uses SIF results for profile learning in CBF. CBF learns the importance of terms through using both SIF's predicted ratings and a

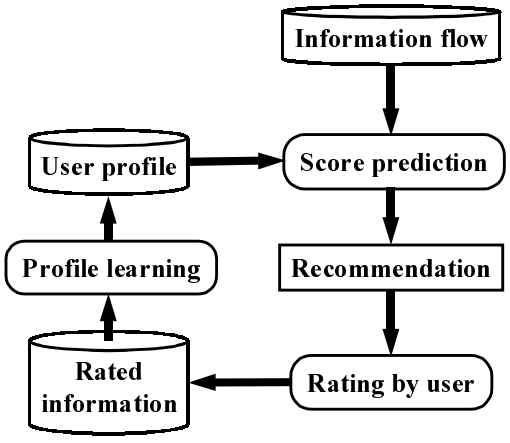

Fig. 1: CBF Configuration

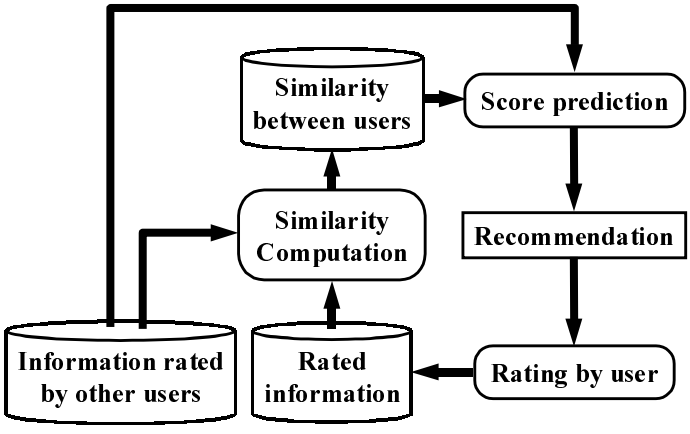

Fig. 2: SIF Configuration

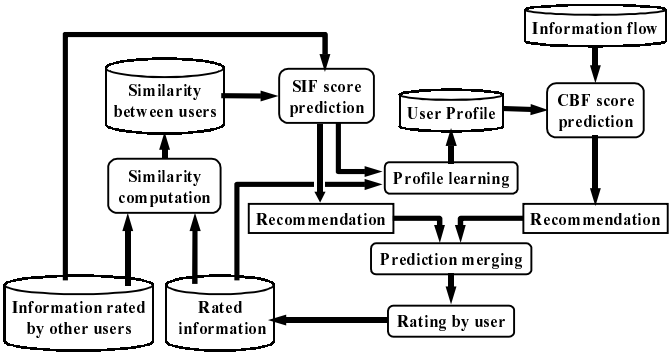

Fig. 3: Combination Method Configuration

user's real ratings.

The feature combination method, for example [Basu and Hirsh, 1997] integrates content features and social features before filtering. The rating data of a specific content is added to the content features (in this paper, called the document vector). A filtering method is then applied to the integrated data.

3. Improvement by Reliability

The goal of combination methods is to develop a method that has the advantages of both CBF and SIF (section 2.3). The feature combination method has problem III, unlearned terms. In contrast, the prediction combination method makes up for the weaknesses of each filtering method, specifically: I) the values of figures and tables and II) new information found by the prediction merging. It also uses strengthened learning to reduce problem III, unlearned terms.

However, the prediction combination method needs to be investigated further. For example, if there are few users who have rated specific information, CBF prediction of the information is more reliable than SIF prediction. In addition, if only predictions with high reliabilities are used to strengthen the learning in CBF, more accurately recommendations can be generated.

Therefore, in this paper, a new combination method, a "hybrid method," is proposed that introduces the reliability of predictions of CBF and SIF into the prediction combination method. The reliability of the predictions is affected by the amount of data that is used in the prediction. In CBF, a prediction score is calculated with a document vector and a user profile. In SIF, user similarities and other users' scores are used for calculating predictions. Thus, in the hybrid method, the reliability of predictions is estimated based on the quantity of the data used in the prediction. In the strengthened learning stage, only predictions that have high reliability are added. In the prediction merging stage, the weights of CBF and SIF are adjusted based on their reliabilities, and these are then used to generate the final prediction.

4. Experimental Evaluation

To evaluate the efficiency of the hybrid method, the prediction performance of both strengthened learning and prediction merging is measured. For strengthened learning, three experiments of without-strengthen, simple strengthen and strengthen with reliability were carried out. For the prediction merging, simple, switch and blend merging were examined.

4.1 Experimental Data

To obtain rating data, we ran a patent clipping service in our laboratory. In this service, user ratings were recorded as integer numbers in the range of 1 to 5. Larger scores indicate that the user had a higher interest in the content.

In these experiments, there were 45 users, all of whom used the service often. These users provided 10,709 ratings for 3,000 different patents. Table 1 is a histogram of the ratings. In this test, document vectors were generated to extract nouns (other than numerals) from the titles and the summaries of each patent. Table 2 shows the number of terms in the document vectors.

4.2 Methodology of the Experiment

To compare each method and to decide the value for each parameter, a 10-fold cross-validation was used. In the 10-fold cross-validation, the rating data was divided into 10 subsets. Each subset was left out from the rating data, and predicted from the other nine subsets using each method. As a result, predicted scores could be generated for each real score. Finally, the prediction performance was measured by comparing the predicted scores to the real scores.

Performance measures: The prediction performance was measured by both precision and cut-off rank. Precision is defined as the proportion of highly rated information in a recommended information set. This precision represented the service quality that a user can observe directly. In the experiments, scores of 5 and 4 were defined as highly rated information. In addition, the cut-off rank was defined as the threshold for determining the amount of information to be recommended. Information that had predicted scores above the cut-off rank was put on the recommendation list.

4.3 Preliminary Experiments

Before measuring performance, preliminary experiments were carried out to decide the value of parameters and a reliability (error) estimation expression.

Parameters: In experiments, CBF and SIF used the al-

Table 1: Histogram of Ratings

Sum	Ratings				
	1	2	3	4	5
10,709	5,165	1,997	1,547	1,270	730

Table 2: Number of Terms

Documents	Terms		
	Mean	Max	Min
3,000	32.58	66	10

gorithms explained in section 2. There were three parameters: α for the learning speed of CBF, β for the similarity threshold of SIF, and γ for the switching threshold of the prediction merging (see 4.4.2). Values that gave the best average precision were chosen for these parameters. Average precision was taken from the range that includes the top of the cut-off to 2000th ranking information.

Reliability Estimation: The measure of reliability used in this paper is the error value, which is defined by the square of the difference between the predicted rating and the user's true rating. In this experiment, errors were estimated with polynomial expressions based on the amount of data used in the prediction. In CBF, a prediction score is calculated by using a document vector and a user profile; and in SIF, it is calculated by using both user similarities and other users' scores.

In preliminary experiments about CBF, the best estimation was given by using the quantity of terms in a particular profile as the amount of the profile. It was also found that the best estimate was given by using the quantity of terms that appear in both the document vector and the profile as the amount of the document vector. In preliminary experiments for SIF, the best estimation was given by using the sum of the user similarities among similar users as the number of similar users.

For the estimating expression, linear, quadratic and cubic polynomial expressions were tried. It was found that the best estimation was given by using the quadratic expression that included the amount of data and the reciprocal numbers of the amount. The parameters of the expression were calculated by the least-squares method using the results of the10-fold cross-validation.

The order of the deciding parameters was as follows: first the SIF α, then the SIF parameters for error estimation, then the CBF β, then the CBF parameters for estimating error, and finally the γ of the prediction merging.

Ideally, it is better to use data for deciding parameters that is different than that data used to measure filtering performance. This study used the same data for deciding the parameters as for measuring performance, because there was not enough data available. However, the target data was not included in the source data used for predicting the scores of the target data, since the 10-fold cross-validation is used in the experiments.

4.4 Summary of Results

These methods were then compared in the graphs of precision vs. the cut-off rank. In the graphs, the horizontal axes represent the cut-off ranks, and the vertical axes represent average precision. Table 1 shows there were 730 recommendations with a score of 5, and 2,000 recommendations with scores of 4 or 5. Accordingly, the cut-off rank in an actual recommendation service should be 700~2000. Thus, the precision in this range is important to evaluate methods.

4.4.1 Experiment 1: Strengthened Learning

The first experiment compared three strengthened learning methods: without-strengthen, simple strengthen and strengthen with reliability. **Simple strengthen** chooses higher predicted scores N/2 and lower predicted scores N/2. **Strengthen with reliability** chooses higher predicted scores N and lower predicted scores N. Then, based on the 2N of the predicted scores chosen, the method chooses highly reliable scores N. N is the number of scores rated by the current target user.

Simple strengthen and strengthen with reliability methods use predicted scores as much as it uses real score ratings, because a predicted score contains some noise. If CBF learns too much with a predicted score, noise may affect the precision. Besides this, these methods used predicted scores that had both higher and lower ratings, because user interest can be learned with an extreme score better than with a middle score.

To compare the performances of different strengthened learning methods, the precision of CBF was measured for each of these three methods. Figure 4 shows the results of the first experiment. The results show that strengthen with reliability gave the best precision.

4.4.2 Experiment 2: Prediction Merging

The second experiment compares three prediction merging methods: simple merging, switch merging and blend merging. **Simple merging** does not consider estimated error. Basically, it chooses an SIF prediction if one exists. It only chooses a CBF prediction if SIF cannot predict a score. **Switch merging** chooses a prediction that has a low estimated error. Basically, if the value of (the estimated error of SIF/the estimated error of CBF) is less than the threshold γ, the predicted score of SIF is chosen. If it is greater than the threshold γ, the CBF prediction is chosen. **Blend merging** blends the CBF prediction and the SIF prediction using errors as weights. Basically, the final prediction is calculated with the following expressions: (predicted score of SIF * estimated error of CBF + predicted score of CBF * estimated error of SIF) / (estimated error of CBF + estimated error of SIF)

Figure 5 shows the results of the second experiment. Each prediction merging method used strengthened learning with reliability. The old combination method used simple strengthen and simple merging. The switch merging showed the best results for cut-off ranks below 1,600. The old combination showed the best results over this rank, although differences of precision between the best methods and the second-best methods (switch merging) were very small.

To verify that these differences are significant, future experiments are planed which involve more users and rating data than the experiments presented in this paper.

The error-estimating expressions are determined to be heuristic. Recently, theoretical combinations of predictions have been researched (ex. [Blum and Mitchell, 1998]). In future research, the author will include a theo-

retical analysis of error estimation and prediction merging.

5. Conclusions

In this paper, a hybrid method is proposed by introducing the reliability of predictions into a combination of CBF and SIF. There are two stages in this combination. In the strengthened learning stage, CBF learns a user profile through using both user's real ratings and SIF's predicted ratings that have high reliabilities. The prediction merging stage integrates the predictions of CBF and SIF. Weights of CBF and SIF in the final prediction are adjusted based on reliability. Besides, the reliability of predictions is estimated based on the quantity of data used in the prediction.

The results of experiments showed that the combination of switch merging and strengthen with reliability works best in high cut-off ranks. While, the combination of simple merging and simple strengthen works best in lower ranks.

The experimental results also showed that SIF contributes to strengthen learning in CBF. Balabanovic and Shoham, (1997) report that SIF needs more than a few hundred users to ensure good prediction. In our experiments, there were only 45 users. The hybrid method showed excellent performance, although the filtering performance of SIF alone was not ideal.

Acknowledgments

A part of this research and development is funded by the Next-Generation Digital Library Project under the auspices of the Japan Information Processing Development Center (JIPDEC).

References

[Ariyoshi and Ichiyama, 1997] Yusuke Ariyoshi and Shunji Ichiyama. An Information Filtering Method Combined Content Based Filtering and Social Information Filtering. *Proceedings of IEICE 8th Workshop on Data Engineering*, pages49—54, Kobe, Japan, March 1997. The Institute of Electronics, Information and Communication Engineers.

[Balabanovic and Shoham, 1997] Marko Balabanovic and Yoav Shoham. Fab: Content Based, Collaborative Recommendation. *Communications of the ACM*, 40(3):66--72, March 1997.

[Basu and Hirsh, 1997] Chumki Basu and Haym Hirsh. Learning to Recommend. *IJCAI-97 Poster Session Abstracts*, page13, Nagoya, Japan, August 1997. International Joint Committee on Artificial Intelligence.

[Billsus and Pazzani, 1998] Daniel Billsus and Michael J. Pazzani. Learning Collaborative Information Filter. *Proceedings of the 15th International Conference on Machine Learning*, page46--54, Madison, Wisconsin, July 1998. Morgan Kaufmann Publishers.

[Nakamura *et al.*, 1997] Atsuyoshi Nakamura, Naoki Abe, Hiroshi Mamitsuka, and Hiroyasu Toba. Learning Personal Preferences by On-line Prediction Algorithms. *IJCAI-97 Poster Session Abstracts*, page23, Nagoya, Japan, August 1997. International Joint Committee on Artificial Intelligence.

[Shardanand and Maes, 1995] Upendra Shardanand and Pattice Maes. Social Information Filtering: Algorithms for Automating "Word of Mouth". *Proceedings of the Conference of Human Factors in Computing Systems*, pages210--217, Denver, Colorado, May 1995. ACM Press.

[Blum and Mitchell, 1998] Avrim Blum and Tom Mitchell. Combining Labeled and Unlabeled Data with Co-Training. *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, page92--100, Madison, Wisconsin, August 1998. ACM Press.

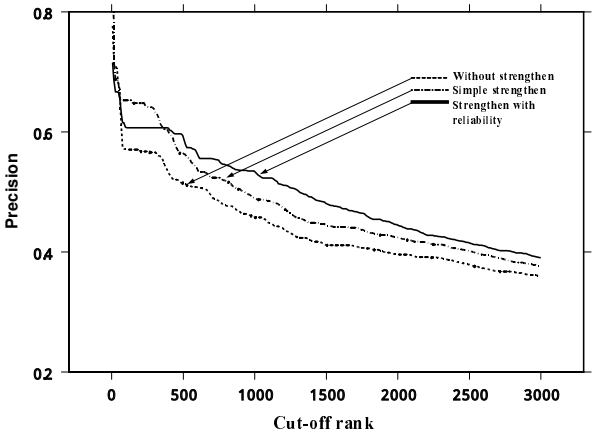

Fig.4: Strengthened Learning

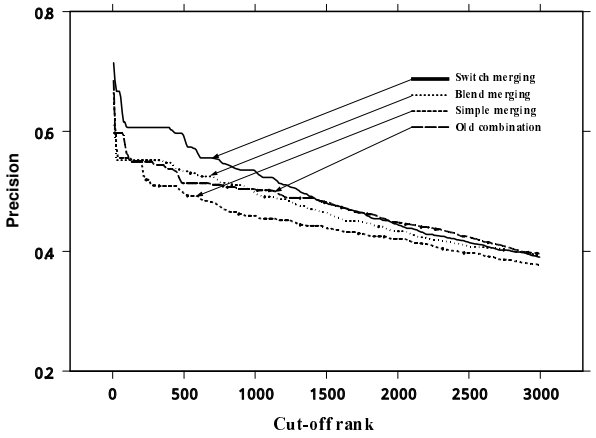

Fig.5: Prediction Merging