

# Enhancement of the Neutrality in Recommendation

Toshihiro Kamishima, Shotaro Akaho,  
and Hideki Asoh  
National Institute of Advanced Industrial Science  
and Technology (AIST)  
AIST Tsukuba Central 2, Umezono 1-1-1,  
Tsukuba, Ibaraki, 305-8568 Japan  
mail@kamishima.net,  
s.akaho@aist.go.jp, h.asoh@aist.go.jp

Jun Sakuma  
University of Tsukuba  
1-1-1 Tennodai, Tsukuba, 305-8577 Japan; and  
Japan Science and Technology Agency  
4-1-8, Honcho, Kawaguchi, Saitama, 332-0012  
Japan  
jun@cs.tsukuba.ac.jp

## ABSTRACT

This paper proposes an algorithm for making recommendation so that the neutrality toward the viewpoint specified by a user is enhanced. This algorithm is useful for avoiding to make decisions based on biased information. Such a problem is pointed out as the filter bubble, which is the influence in social decisions biased by a personalization technology. To provide such a recommendation, we assume that a user specifies a viewpoint toward which the user want to enforce the neutrality, because recommendation that is neutral from any information is no longer recommendation. Given such a target viewpoint, we implemented information neutral recommendation algorithm by introducing a penalty term to enforce the statistical independence between the target viewpoint and a preference score. We empirically show that our algorithm enhances the independence toward the specified viewpoint by and then demonstrate how sets of recommended items are changed.

## Categories and Subject Descriptors

H.3.3 [INFORMATION SEARCH AND RETRIEVAL]:  
Information filtering

## Keywords

neutrality, fairness, filter bubble, collaborative filtering, matrix decomposition, information theory

## 1. INTRODUCTION

A recommender system searches for items and information that would be useful to a user based on the user's behaviors or the features of candidate items [21, 2]. GroupLens [19] and many other recommender systems emerged in the mid-1990s, and further experimental and practical systems have been developed during the explosion of Internet merchandizing. In the past decade, such recommender systems have been introduced and managed at many e-commerce sites to promote items sold at these sites.

Paper presented at the 2012 Decisions@RecSys workshop in conjunction with the 6th ACM conference on Recommender Systems. Copyright © 2012 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

2012 Decisions@RecSys workshop September 9, 2012, Dublin, Ireland

The influence of personalization technologies such as recommender systems or personalized search engines on people's decision making is getting stronger and stronger. For example, at a shopping site, if a customer checks a recommendation list and finds a five-star-rated item, he/she would more seriously consider buying the highly rated item. The filter bubble, which is the selection of the appropriate diversity of information provided to users, is one of problems accompanying the growing influence of recommendation or personalization.

The problem of the filter bubble was recently posed by Pariser [17]. Via the influence of personalized technologies, the topics of information provided to users are becoming restricted to those originally preferred by them, and this restriction is not notified by users. This situation is compared to shutting up each individual in a separate *bubble*. Pariser claimed that due to the obstruction of these bubbles, users lose the opportunity of finding new topics and that sharing public matters throughout our society is getting harder. To discuss this filter bubble problem, a panel discussion was held at the RecSys 2011 conference [20].

During the RecSys panel discussion, panelists made the following assertions about the filter bubble problem. Biased topics would be certainly selected by the influence of personalization, but at the same time, it would be intrinsically impossible to make recommendations that are absolutely neutral from any viewpoint, and the diversity of provided topics intrinsically has a trade-off relation to the fitness of these topics for users' interests or needs. To recommend something, or more generally to select something, one must consider the specific aspect of a thing and must ignore the other aspects of the thing. The panelists also pointed out that current recommender systems fail to satisfy users' information need that they search for a wide variety of topics in the long term.

To solve this problem, we propose an information neutral recommender system that guarantees the neutrality of recommendation results. As pointed out during the RecSys 2011 panel discussion, because it is impossible to make a recommendation that is absolutely neutral from all viewpoints, we consider neutrality from the viewpoint or information specified by a user. For example, users can specify a feature of an item, such as a brand, or a user feature, such as a gender or an age, as a viewpoint. An information neu-

tral recommender system is designed so that these specified features will not affect recommendation results. This system can also be used to avoid the use of information that is restricted by law or regulation. For example, the use of some information is prohibited for the purpose of making recommendation by privacy policies.

We borrowed the idea of fairness-aware mining, which we proposed earlier [11], to build this information neutral recommender system. To enhance the neutrality or the independence in recommendation, we introduce a regularization term that represents the mutual information between a recommendation result and the specified viewpoint.

Our contributions are as follows. First, we present a definition of neutrality in recommendation based on the consideration of why it is impossible to achieve an absolutely neutral recommendation. Second, we propose a method to enhance the neutrality that we defined and combine it with a latent factor recommendation model. Finally, we demonstrate that the neutrality of recommendation can be enhanced and how recommendation results change by enhancing the neutrality.

In section 2, we discuss the filter bubble problem and neutrality in recommendation and define the goal of an information neutral recommender task. An information neutral recommender system is proposed in section 3, and its experimental results are shown in section 4. Sections 5 and 6 cover related work and our conclusion, respectively.

## 2. INFORMATION NEUTRALITY

In this section, we discuss information neutrality in recommendation based on the considerations on the filter bubble problem and the ugly duckling theorem.

### 2.1 The Filter Bubble Problem

We here summarize the filter bubble problem posed by Pariser and the discussion in the panel about this problem at the RecSys 2011 conference. The *filter bubble* problem is a concern that personalization technologies, including recommender systems, narrow and bias the topics of information provided to people while they don't notice these facts [17].

Pariser demonstrated the following examples in a TED talk about this problem [16]. In a social network service, Facebook<sup>1</sup>, users have to specify a group of *friends* with whom they can chat or have private discussions. To help users find their friends, the service has a function to list other users' accounts that are expected to be related to a user. When Pariser started to use Facebook, the system showed a recommendation list that consisted of both conservative and progressive people. However, because he has more frequently selected progressive people as friends, conservative people have been excluded from his recommendation list by a personalization functionality. Pariser claimed that the system excluded conservative people without his permission and that he lost the opportunity of getting a wide variety of opinions.

He furthermore demonstrated a collection of search results

<sup>1</sup><http://www.facebook.com>

from Google<sup>2</sup> for the query "Egypt" during the Egyptian uprising in 2011 from various people. Even though such a highly important event was occurring, only sightseeing pages were listed for some users instead of news pages about the Egyptian uprising, due to the influence of personalization. In this example, he claimed that personalization technology spoiled the opportunity to obtain information that should be commonly shared in our society.

We consider that Pariser's claims can be summarized as follows. The first point is the problem that users lost opportunities to obtain information about a wide variety of topics. A chance to know things that could make users' lives fruitful was lessened. The second point is the problem that each individual obtains information that is too personalized, and thus the amount of shared information is decreased. Pariser claimed that the loss of sharing information is a serious obstacle for building consensus in our society. He claimed that the loss of the ability to share information is a serious obstacle for building consensus in our society.

RecSys 2011, which is a conference on recommender systems, held a panel discussion the topic of which was this filter bubble problem [20]. This panel concentrated on the following three arguing points. (1) Are there *filter bubbles*? Resnick pointed out the possibility that personalization technologies narrow users' experience in the mid 1990s. Because selecting specific information by definition leads to ignoring other information, the diversity of users' experiences intrinsically have a trade-off relation to the fitness of information for users' interests. As seen in the difference between the perspective of al-Jazeera and that of Fox News, this problem exists irrespective of personalization. Further, given signals or expressions of users' interest, it is difficult to adjust how much a system should meet those interests.

(2) To what degree is personalized filtering a problem? There is no absolutely neutral viewpoint. On the other hand, the use of personalized filtering is inevitable, because it is not feasible to exhaustively access the vast amount of information in the universe. One potential concern is the effect of selective exposure, which is the tendency to get reinforcement of what people already believe. According to the results of studies about this concern, this is not so serious, because people viewing extreme sites spend more time on mainstream news as well.

(3) What should we as a community do to address the filter bubble issue? To adjust the trade-off between diversity and fitness of information, a system should consider users' immediate needs as well as their long-term needs. Instead of selecting individual items separately, a recommendation list or portfolio should be optimized as a whole.

### 2.2 The Neutrality in Recommendation

The absence of the absolutely neutral viewpoint is pointed out in the above panel. We here more formally discuss this point based on the ugly duckling theorem.

The *ugly duckling theorem* is a classical theorem in pattern recognition literature that asserts the impossibility of classi-

<sup>2</sup><http://www.google.com>

fication without weighing certain features or aspects of objects against the others [26]. Consider a case that  $n$  ducklings are represented by at least  $\log_2 n$  binary features, for example, black feathers or a fat body, and are classified into positive or negative classes based on these features. If the positive class is represented by Boolean functions of binary features, it is easy to prove that the number of possible functions that classify an arbitrary pair of ducklings into a positive class is  $2^{n-2}$ , even if choosing any pairs of ducklings. Provided that the similarity between a pair of ducklings is measured by the number of functions that classify them into the same class, the similarity between an ugly duckling and an arbitrary normal duckling is equal to the similarity between any pair of ducklings. In other words, an ugly duckling looks like a normal duckling.

Why is an ugly duckling ugly? As described above, an ugly duckling is as ugly as a normal duckling, if all features and functions are treated equally. The attention to an arbitrary feature such as black feathers makes an ugly duckling ugly. When we classify something, we of necessity pay attention to certain features, aspects, or viewpoints of classified objects. Because recommendation is considered as a task to classify items into a relevant class or an irrelevant one, certain features or viewpoints must be inevitably weighed when making recommendation. Consequently, the absolutely neutral recommendation is intrinsically impossible.

We propose a neutral recommendation task other than the absolutely neutral recommendation. Recalling the ugly duckling theorem, we must focus on certain features or viewpoints in classification. This fact indicates that it is feasible to make a recommendation that is neutral from a specific viewpoint instead of all viewpoints. We hence advocate an *Information Neutral Recommender System* that enhances the neutrality in recommendation from the viewpoint specified by a user. In the case of Pariser’s Facebook example, a system enhances the neutrality so that recommended friends are conservative or progressive, but the system is allowed to make biased decisions in terms of the other viewpoints, for example, the birthplace or age of friends.

### 3. AN INFORMATION NEUTRAL RECOMMENDER SYSTEM

In this section, we formalize a task of information neutral recommendation and show a solution algorithm for this task.

#### 3.1 Task Formalization

In [8], recommendation tasks are classified into *Recommending Good Items* that meet a user’s interest, *Optimizing Utility* of users, and *Predicting Ratings* of items for a user. Among these tasks, we here concentrate on the task of predicting ratings.

We formalize an information neutral variant of a predicting ratings task.  $x \in \{1, \dots, n\}$  and  $y \in \{1, \dots, m\}$  denote a user and an item, respectively. An event  $(x, y)$  is a pair of a specific user  $x$  and a specific item  $y$ . Here,  $s$  denotes a rating value of  $y$  as given by  $x$ . We here assume that the domain of ratings is real values, though domain of ratings is commonly a set of discrete values, e.g.,  $\{1, \dots, 5\}$ . These variables are common for an original predicting ratings task.

To treat the information neutrality in recommendation, we additionally introduce a viewpoint variable,  $v$ , which indicates a viewpoint neutrality from which is enhanced. This variable is specified by a user, and its value depends on an event. Possible examples of a viewpoint variable are a user’s gender, which depends on a user part of an event, movie’s release year, which depends on an item’s part of an event, and a timestamp when a user rates an item, which depends on both elements in an event. In this paper, we restrict the domain of a viewpoint variable to a binary type, 0, 1, but it is easy to extend to a multinomial case. An example consists of an event,  $(x, y)$ , a rating value for the event,  $s$ , and a viewpoint value for the event,  $v$ . A training set is a set of  $N$  examples,  $\mathcal{D} = \{(x_i, y_i, s_i, v_i)\}$ ,  $i = 1, \dots, N$ .

Given a new event,  $(x, y)$ , and its corresponding viewpoint value,  $v$ , a rating prediction function,  $\hat{s}(x, y, v)$ , predicts a rating value of an item  $y$  by a user  $x$ . While this rating prediction function is estimated in our task setting, a loss function,  $\text{loss}(s^*, \hat{s})$ , and a neutrality function,  $\text{neutral}(\hat{s}, v)$ , are given as task inputs. A loss function represents the dissimilarity between a true rating value,  $s^*$ , and a predicted rating value,  $\hat{s}$ . A neutrality function quantifies the degree of the neutrality of a rating value from a viewpoint expressed by a viewpoint variable. Given a training set,  $\mathcal{D}$ , a goal of an information neutral recommendation (predicting rating case) is to acquire a rating prediction function,  $\hat{s}(x, y, v)$ , so that the expected value of a loss function is as small as possible and the expected value of a neutral function is as large as possible over  $(x, y, v)$ . We formulate this goal by finding a rating prediction function,  $\hat{s}$ , so as to minimize the following objective function:

$$\text{loss}(s^*, \hat{s}(x, y, v)) - \eta \text{neutral}(\hat{s}(x, y, v), v), \quad (1)$$

where  $\eta > 0$  is a parameter to balance between the loss and the neutrality.

#### 3.2 A Prediction Model

In this paper, we adopt a latent factor model for predicting ratings. This latent factor model, which is a kind of a matrix decomposition model, is defined as equation (3) in [12], as follows:

$$\hat{s}(x, y) = \mu + b_x + c_y + \mathbf{p}_x \mathbf{q}_y^\top, \quad (2)$$

where  $\mu$ ,  $b_x$ , and  $c_y$  are global, per user, and per item bias parameters, respectively, and  $\mathbf{p}_x$  and  $\mathbf{q}_y$  are  $K$ -dimensional parameter vectors, which represent the cross effects between users and items. We adopt a squared loss as a loss function. As a result, parameters of a rating prediction function can be estimated by minimizing the following objective function:

$$\sum_{(x_i, y_i, s_i) \in \mathcal{D}} (s_i - \hat{s}(x_i, y_i))^2 + \lambda R, \quad (3)$$

where  $R$  represents an  $L_2$  regularizer for parameters  $b_x$ ,  $c_y$ ,  $\mathbf{p}_x$ , and  $\mathbf{q}_y$ , and  $\lambda$  is a regularization parameter. Once we learned the parameters of a rating prediction function, we can predict a rating value for any event by applying equation (2).

We then extend this model to enhance the information neutrality. First, we modify the model of equation (2) so as to depend on the value of a viewpoint variable,  $v$ . For each value of  $v$ , 0 and 1, we prepare a parameter set,  $\mu^{(v)}$ ,  $b_x^{(v)}$ ,

$c_y^{(v)}$ ,  $\mathbf{p}_x^{(v)}$ , and  $\mathbf{q}_y^{(v)}$ . One of parameter sets is chosen according as a value of  $v$ , and we get a rating prediction function:

$$\hat{s}(x, y, v) = \mu^{(v)} + b_x^{(v)} + c_y^{(v)} + \mathbf{p}_x^{(v)} \mathbf{q}_y^{(v)\top}. \quad (4)$$

We next define a neutrality function to quantify the degree of the information neutrality from a viewpoint variable,  $v$ . In this paper, we borrow an idea from [11] and quantify the degree of the information neutrality by negative mutual information under the assumption that neutrality is regarded as statistical independence. A neutrality function is defined as:

$$\begin{aligned} -I(\hat{s}; v) &= \sum_{v \in \{0,1\}} \int \Pr[\hat{s}, v] \log \frac{\Pr[\hat{s}|v]}{\Pr[\hat{s}]} d\hat{s} \\ &= \sum_{v \in \{0,1\}} \Pr[v] \int \Pr[\hat{s}|v] \log \frac{\Pr[\hat{s}|v]}{\Pr[\hat{s}]} d\hat{s}. \end{aligned} \quad (5)$$

The marginalization over  $v$  is then replaced with the sample mean over a training set,  $\mathcal{D}$ , and we get

$$\frac{1}{N} \sum_{(v) \in \mathcal{D}} \int \Pr[\hat{s}|v] \log \frac{\Pr[\hat{s}|v]}{\Pr[\hat{s}]} d\hat{s}. \quad (6)$$

Note that  $\Pr[\hat{s}]$  can be calculated by  $\sum_v \Pr[\hat{s}|v] \Pr[v]$ , and we use a sample mass function as  $\Pr[v]$ .

Now, all that we have to do is to compute distribution  $\Pr[\hat{s}|v]$ , but this computation is difficult. This is because a value of a function  $\hat{s}$  is not probabilistic but rather deterministic depending on  $x$ ,  $y$ , and  $v$ ; and thus distribution  $\Pr[\hat{s}|x, y, v]$  has a form of collection of Dirac's delta functions,  $\delta(\hat{s}(x, y, v))$ .  $\Pr[\hat{s}|v]$  can be obtained by marginalizing this distribution over  $x$  and  $y$ . As a result,  $\Pr[\hat{s}|v]$  also becomes a hyper function like  $\Pr[\hat{s}|x, y, v]$ , and it is not easy to manipulate. We therefore introduce a histogram model to represent  $\Pr[\hat{s}|v]$ . Values of predicted ratings,  $\hat{s}$ , are divided into bins, because sample ratings are generally discrete. The distribution  $\Pr[\hat{s}|v]$  is expressed by a histogram model. By replacing  $\Pr[\hat{s}|v]$  with  $\tilde{\Pr}[\hat{s}|v]$ , equation (6) becomes

$$\frac{1}{N} \sum_{(v) \in \mathcal{D}} \sum_{\hat{s} \in \text{Bin}} \tilde{\Pr}[\hat{s}|v] \log \frac{\tilde{\Pr}[\hat{s}|v]}{\tilde{\Pr}[\hat{s}]}, \quad (7)$$

where Bin denotes a set of bins of a histogram. Note that because a distribution function,  $\Pr[\hat{s}|v]$ , is replaced with a probability mass function,  $\tilde{\Pr}[\hat{s}|v]$ , an integration over  $\hat{s}$  is replaced with the summation over bins.

By substituting equations (4) and (7) into equation (1) and adopting a squared loss function as in the original latent factor case, we obtain an objective function of an information neutral recommendation model:

$$\mathcal{L}(\mathcal{D}) = \sum_{(x_i, y_i, s_i, v_i) \in \mathcal{D}} (s_i - \hat{s}(x_i, y_i, v_i))^2 + \eta I(\hat{s}; v) + \lambda R, \quad (8)$$

where a regularization term,  $R$ , is a sum of  $L_2$  regularizers of parameter sets for each value of  $v$ . Model parameters,  $\{\mu^{(v)}, b_x^{(v)}, c_y^{(v)}, \mathbf{p}_x^{(v)}, \mathbf{q}_y^{(v)}\}$ ,  $v \in \{0, 1\}$ , are estimated so as to minimize this objective function. However, it is very difficult to derive an analytical form of gradients of this objective function, because the histogram transformation used for expressing  $\Pr[\hat{s}|v]$  is too complicated. We therefore adopt the

Powell optimization method, because it can be applied without computing gradients.

## 4. EXPERIMENTS

We implemented our information neutral recommender system in the previous section and applied it to a benchmark data set.

### 4.1 A Data Set

We used a MovieLens 100k data set [7] in our experiments. As described in section 3.2, we adopted the Powell method for optimizing an objective function. Unfortunately, this method is too slow to apply to a large data set, because the number of evaluation times of an objective function becomes very large to avoid the computation of gradients. Therefore, we shrank the MovieLens data set by extracting events whose user ID and item ID were less than or equal to 200 and 300, respectively. This shrunken data set contained 9,409 events, 200 users, and 300 items.

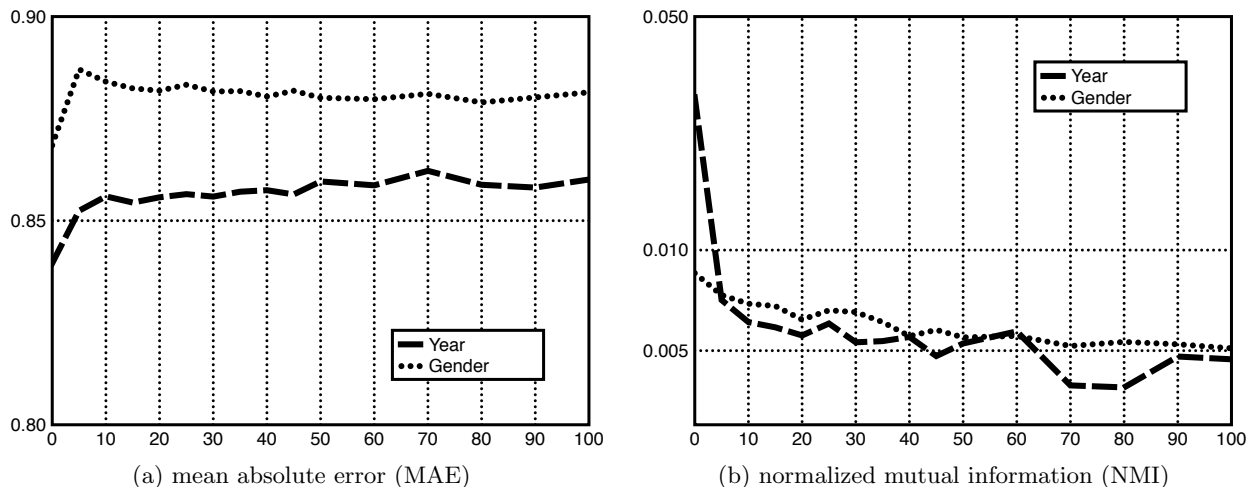
We tested the following two types of viewpoint variable. The first type of variable, **Year**, represents whether a movie's release year is newer than 1990, which depends on an item part of an event. In [12], Koren reported that the older movies have a tendency to be rated higher, perhaps because only masterpieces have survived. When adopting **Year** as a viewpoint variable, our recommender enhances the neutrality from this masterpiece bias. The second type of variable, **Gender**, represents the user's gender, which depends on the user part of an event. The movie rating would depend on the user's gender, and our recommender enhances the neutrality from this factor.

### 4.2 Experimental Conditions

We used the implementation of the Powell method in the SciPy package [22] as an optimizer for an objective function (8). To initialize parameter, events in a training set,  $\mathcal{D}$ , were first divided into two sets according to their viewpoint values. For each value of a viewpoint variable, parameters are initialized by minimizing an objective function of an original latent factor model (equation (3)). For the convenience in implementation, a loss term of an objective was scaled by dividing it by the number of training examples, and an  $L_2$  regularizer was scaled by dividing it by the number of parameters. We use a regularization parameter  $\lambda = 0.01$  and the number of latent factors,  $K = 1$ , which are the lengths of vectors  $\mathbf{p}^{(v)}$  or  $\mathbf{q}^{(v)}$ . Because the original rating values are 1, 2, ..., 5, we adopted five bins whose centers are 1, 2, ..., 5, in equation (7). We performed a five-fold cross-validation procedure to obtain evaluation indices of the prediction accuracy and the neutrality from a viewpoint variable.

### 4.3 Experimental Results

Experimental results are shown in Figure 1. Figure 1(a) shows the changes of prediction errors measured by a mean absolute error (MAE) index. The smaller value of this index indicates better prediction accuracy. Figure 1(b) shows the changes of the mutual information between predicted ratings and viewpoint values. The smaller mutual information indicates a higher level of neutrality. Mutual information is normalized into the range [0, 1] by the method of employing the geometrical mean in [24]. Note that distribution  $\Pr[\hat{s}|v]$



**Figure 1: Changes of the degrees of neutrality accompanying the increase of a neutrality parameter**

NOTE : Figure 1(a) shows the changes of prediction errors measured by a mean absolute error (MAE) index. The smaller value of this index indicates better prediction accuracy. Figure 1(b) shows the changes of the mutual information between predicted ratings and viewpoint values. The smaller mutual information indicates a higher level of neutrality. The X-axes of these figures represent parameter values of  $\eta$ . Dashed lines and dotted lines show the results using Year and Gender as viewpoint variables, respectively.

is required to compute this mutual information, and we used the same histogram model as in equation (7). The X-axes of these figures represent parameter values of  $\eta$ , which balance the prediction accuracy and the neutrality. These parameters were changed from 0, at which the neutrality term was completely ignored, to 100, at which the neutrality was highly emphasized. Dashed lines and dotted lines show the results using Year and Gender as viewpoint variables, respectively.

MAE was 0.90, when offering a mean score, 3.74, for all users and all items. In Figure 1(a), MAEs were better than this baseline, which is perfectly neutral from all viewpoints. Furthermore, the increase of MAEs as the neutrality parameter,  $\eta$ , was not so serious. Turning to the Figure 1(b), this demonstrates that the neutrality is enhanced as the neutrality parameter,  $\eta$ , increases from both viewpoints, Year and Gender. By drawing attention to the fact that the Y-axis is logarithmic, we can conclude that an information neutrality term is highly effective. In summary, our information neutral recommender system successfully enhanced the neutrality without seriously sacrificing the prediction accuracy.

Figure 2 shows the changes of mean predicted scores. In both figures, the X-axes represent parameter values of  $\eta$ , and the Y-axes represent mean predicted scores for each case of using different viewpoint value. Figure 1(a) shows mean predicted scores when a viewpoint variable is Year. Dashed and dotted lines show the results under the condition a viewpoint variable is “before 1990” and “after 1991”, respectively. Figure 1(b) shows mean predicted scores when a viewpoint variable is Gender. Dashed and dotted lines show the results obtained by setting a viewpoint to “male” and “female”, respectively.

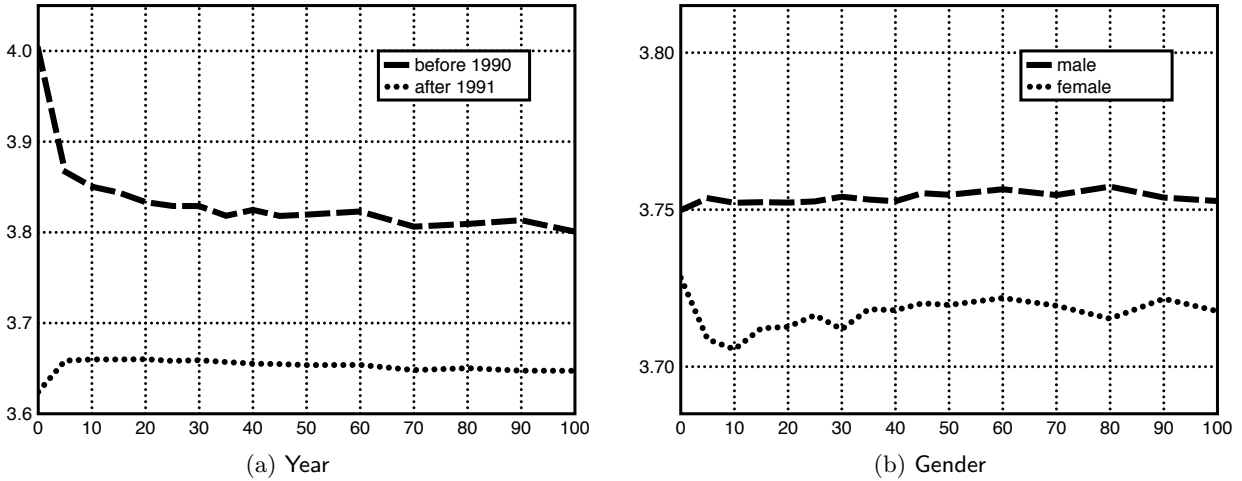
We first discuss a case that a viewpoint variable is Year. According to Figure 1(b), neutrality was drastically improved

in the interval that  $\eta$  is between 0 and 10. By observing the corresponding interval in Figure 2, two lines that were obtained for different viewpoints became close each other. This means that prediction scores become less affected by a viewpoint value, and this corresponds the improvement of neutrality. After this range, the decrease of NMI became smaller in Figure 1(b), and the lines in the corresponding interval in Figure 2 were nearly parallel. This indicated that the difference between two score sequences less changes, and the improvement in neutrality did too. We move on to a Gender case. By comparing the changes of NMI between Year and Gender cases in Figure 1(b), the decrease of NMI in a Gender case was much smaller than that of a Year case. This phenomenon could be confirmed by the fact that two lines were nearly parallel in Figure 2(b). This is probably because the score differences in a Gender case are much smaller than those in a Year at the point  $\eta = 0$ , and there is less margin for improvement. Further investigation will be required in this point.

## 5. RELATED WORK

To enhance the neutrality, we borrowed an idea from our previous work [11], which is an analysis technique for fairness/discrimination-aware mining. Fairness/discrimination-aware mining is a general term for mining techniques designed so that sensitive information does not influence mining results. In [18], Pedreschi et al. first advocated such mining techniques, which emphasized the unfairness in association rules whose consequents include serious determinations. Like this work, a few techniques for detecting unfair treatments in mining results have been proposed [14, 25]. These techniques might be useful for detecting biases in recommendation.

Another type of fairness-aware mining technique focuses on classification designed so that the influence of sensitive information to classification results is reduced [11, 3, 10] These



**Figure 2: Changes of mean predicted scores accompanying the increase of a neutrality parameter**

NOTE : In both figures, the X-axes represent parameter values of  $\eta$ , and the Y-axes represent mean predicted scores for each case of different viewpoint value. Figure 1(a) shows mean predicted scores when a viewpoint variable is Year. Dashed and dotted lines show the results under the condition a viewpoint variable is “before 1990” and “after 1991”, respectively. Figure 1(b) shows mean predicted scores when a viewpoint variable is Gender. Dashed and dotted lines show the results obtained by setting a viewpoint to “male” and “female”, respectively.

techniques would be directly useful in the development of an information neutral variant of content-based recommender systems, because content-based recommenders can be implemented by adopting classifiers.

Information neutrality can be considered as diversity in recommendation in a broad sense. McNee et al. pointed out the importance of factors other than prediction accuracy, including diversity, in recommendation [15]. Topic diversification is a technique for enhancing the diversity in a recommendation list [27]. Smyth et al. proposed a method for changing the diversity in a recommendation list based on a user’s feedback [23].

There are several reports about the influence of recommendations on the diversity of items accepted by users. Celma et al. reported that recommender systems have a popularity bias such that popular items have a tendency to be recommended more and more frequently [4]. Fleder et al. investigated the relation between recommender systems and their impact on sales diversity by simulation [6]. Levy et al. reported that sales diversity could be slightly enriched by recommending very unpopular items [13].

Because information neutral recommenders can be used to avoid the exploitation of private information, these techniques are related to privacy-preserving data mining [1]. Independent component analysis might be used to maintain the independence between viewpoint values and recommendation results [9]. In a broad sense, information neutral recommenders are a kind of cost-sensitive learning technique [5], because these recommenders are designed to take into account the costs of enhancing the neutrality.

## 6. CONCLUSION

In this paper, we proposed an information neutral recommender system that enhanced neutrality from a viewpoint

specified by a user. This system is useful for alleviating the filter bubble problem, which is a concern that personalization technologies narrow users’ experience. We then developed an information neutral recommendation algorithm by introducing a regularization term that quantifies neutrality by mutual information between a predicted rating and a viewpoint variable expressing a user’s viewpoint. We finally demonstrated that neutrality could be enhanced without sacrificing prediction accuracy by our algorithm.

The most serious issue of our current algorithm is scalability. This is mainly due to the difficulty in deriving the analytical form of gradients of an objective function. We plan to develop another objective function whose gradients can be derived analytically. The degree of statistical independence is currently quantified by mutual information. We want to test other indexes, such as kurtosis, which are used for independent component analysis. We will develop an information neutral version of other recommendation models, such as pLSI/LDA or nearest neighbor models.

## 7. ACKNOWLEDGMENTS

We would like to thank for providing a data set for the GroupLens research lab. This work is supported by MEXT/JSPS KAKENHI Grant Number 16700157, 21500154, 22500142, 23240043, and 24500194, and JST PRESTO 09152492.

## 8. REFERENCES

- [1] C. C. Aggarwal and P. S. Yu, editors. *Privacy-Preserving Data Mining: Models and Algorithms*. Springer, 2008.
- [2] J. Ben Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5:115–153, 2001.
- [3] T. Calders and S. Verwer. Three naive bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, 21:277–292, 2010.

- [4] Ò. Celma and P. Cano. From hits to niches?: or how popular artists can bias music recommendation and discovery. In *Proc. of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, 2008.
- [5] C. Elkan. The foundations of cost-sensitive learning. In *Proc. of the 17th Int'l Joint Conf. on Artificial Intelligence*, pages 973–978, 2001.
- [6] D. Fleder and K. Hosanagar. Recommender systems and their impact on sales diversity. In *ACM Conference on Electronic Commerce*, pages 192–199, 2007.
- [7] GroupLens research lab, university of minnesota. (<http://www.grouplens.org/>).
- [8] A. Gunawardana and G. Shani. A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, 10:2935–2962, 2009.
- [9] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001.
- [10] F. Kamiran, T. Calders, and M. Pechenizkiy. Discrimination aware decision tree learning. In *Proc. of the 10th IEEE Int'l Conf. on Data Mining*, pages 869–874, 2010.
- [11] T. Kamishima, S. Akaho, and J. Sakuma. Fairness-aware learning through regularization approach. In *Proc. of The 3rd IEEE Int'l Workshop on Privacy Aspects of Data Mining*, pages 643–650, 2011.
- [12] Y. Koren. Collaborative filtering with temporal dynamics. In *Proc. of the 15th Int'l Conf. on Knowledge Discovery and Data Mining*, pages 447–455, 2009.
- [13] M. Levy and K. Bosteels. Music recommendation and the long tail. In *WOMRAD 2010: Recsys 2010 Workshop on Music Recommendation and Discovery*, 2010.
- [14] B. T. Luong, S. Ruggieri, and F. Turini. k-NN as an implementation of situation testing for discrimination discovery and prevention. In *Proc. of the 17th Int'l Conf. on Knowledge Discovery and Data Mining*, pages 502–510, 2011.
- [15] S. M. McNee, J. Riedl, and J. A. Konstan. Accurate is not always good: How accuracy metrics have hurt recommender systems. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1097–1101, 2006.
- [16] E. Pariser. The filter bubble. (<http://www.thefilterbubble.com/>).
- [17] E. Pariser. *The Filter Bubble: What The Internet Is Hiding From You*. Viking, 2011.
- [18] D. Pedreschi, S. Ruggieri, and F. Turini. Discrimination-aware data mining. In *Proc. of the 14th Int'l Conf. on Knowledge Discovery and Data Mining*, 2008.
- [19] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of Netnews. In *Proc. of the Conf. on Computer Supported Cooperative Work*, pages 175–186, 1994.
- [20] P. Resnick, J. Konstan, and A. Jameson. Panel on the filter bubble. The 5th ACM conference on Recommender systems, 2011.
- (<http://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/>).
- [21] P. Resnick and H. R. Varian. Recommender systems. *Communications of The ACM*, 40(3):56–58, 1997.
- [22] Scipy.org. (<http://www.scipy.org/>).
- [23] B. Smyth and L. McGinty. The power of suggestion. In *Proc. of the 18th Int'l Joint Conf. on Artificial Intelligence*, pages 127–132, 2003.
- [24] A. Strehl and J. Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.
- [25] I. Žliobaitė, F. Kamiran, and T. Calders. Handling conditional discrimination. In *Proc. of the 11th IEEE Int'l Conf. on Data Mining*, 2011.
- [26] S. Watanabe. *Knowing and Guessing – Quantitative Study of Inference and Information*. John Wiley & Sons, 1969.
- [27] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proc. of the 14th Int'l Conf. on World Wide Web*, pages 22–32, 2005.