Considerations on Recommendation Independence for a Find-Good-Items Task

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

ABSTRACT

This paper examines the notion of recommendation independence, which is a constraint that a recommendation result is independent from specific information. This constraint is useful in ensuring adherence to laws and regulations, fair treatment of content providers, and exclusion of unwanted information. For example, to make a job-matching recommendation socially fair, the matching should be independent of socially sensitive information, such as gender or race. We previously developed several recommenders satisfying recommendation independence, but these were all designed for a predicting-ratings task, whose goal is to predict a score that a user would rate. We here focus on another find-good-items task, which aims to find some items that a user would prefer. In this task, scores representing the degree of preference to items are first predicted, and some items having the largest scores are displayed in the form of a ranked list. We developed a preliminary algorithm for this task through a naive approach, enhancing independence between a preference score and sensitive information. We empirically show that although this algorithm can enhance independence of a preference score, it is not fit for the purpose of enhancing independence in terms of a ranked list. This result indicates the need for inventing a notion of independence that is suitable for use with a ranked list and that is applicable for completing a find-good-items task.

CCS CONCEPTS

• **Information systems → Recommender systems**;

KEYWORDS

fairness, recommendation independence, matrix factorization

1 INTRODUCTION

Recommender systems and other personalization technologies, which help to search for items or information predicted to be useful to a user, have become indispensable tools in support of decision-making. To avoid unfairness or bias in the decisions supported by recommender systems, the influence of specific information should be excluded from the prediction process of recommendation.

In other words, independence between recommendation results and specific information should be maintained in the following situations. First, recommendation services must be managed in adherence to laws and regulations. Sweeny presented an example of dubious advertisement placement that appeared to exhibit racial discrimination [21]. In this case, the selection of personalized advertisements should be rendered independent of racial information. Another concern is the fair treatment of information providers. The Federal Trade Commission has been investigating Google to determine whether the search engine ranks its own services higher than those of competitors [3]. In this case, no deliberate manipulation was found. However, an algorithm that can explicitly exclude information about whether content providers are competitors would be helpful for alleviating users' doubts as well as competitors' doubts about unfair manipulations. Finally, recommendation independence is helpful for excluding the influence of unwanted information. Popularity bias, which is the tendency for frequently consumed items to be recommended more frequently [2], is a well-known drawback of recommenders. If information about popularity could be excluded, users could acquire information free from unwanted popularity bias. In summary, excluding the influence of specific information is helpful for the following purposes: adherence to laws and regulations, fair treatment of content providers, and exclusion of unwanted information.

To fulfill the need for excluding the influence of specific information, we formalized a notion of recommendation independence and developed algorithms to enhance it. For this purpose, we exploited a technique developed for fairness-aware data mining [5, 17], whose goal is to analyze data while taking into account potential issues of fairness. Following the notions proposed in the previous studies, we formally define recommendation independence as statistical independence between a recommendation result and specified information. In addition, we developed an independence-enhanced recommender system (IERS) that could satisfy a constraint of recommendation independence [9]. This IERS is also technically challenging and non-trivial, because while there are many techniques for incorporating new types of information, there are very few trials to exclude unwanted information. We developed two approaches for enhancing recommendation independence. One was a regularization approach, which adopted an objective function with a constraint term for imposing recommendation independence [9, 11, 12]. The other was a model-based approach, which adopted a generative model in which ratings and sensitive features were independent [13].

However, all our previous methods targeted a predicting-ratings task, predicting a score of items that a user would rate, although there are other types of recommendation tasks. One such task is a

find-good-items task, whose goal is to find some items that a user would prefer [4, 8]. To complete this type of task, a system predicts preference scores, which quantify how strongly a target user prefers items, for every candidate item. These items are then displayed to a target user in the form of a ranked list sorted according to the predicted scores.

In this paper, we investigate recommendation independence for this find-good-items task. In the case of a predicting-ratings task, we enhanced independence between a predicted rating and a sensitive feature. However, in the find-good-items case, the notion of independence between a ranked list and a sensitive feature is unclear. We therefore examine a naive approach, treating independence between a preference score used for ranking items and a sensitive feature. We develop a preliminary recommendation method to enhance this type of independence by a regularization approach. By applying this method, we empirically inspect the independence from a preference score or a ranked list.

Our contributions can be summarized as follows.

- We develop a preliminary recommendation method for a find-good-items task through an approach of enhancing the independence of a preference score from a sensitive feature.
- We empirically show that the independence of a preference score could be enhanced without sacrificing prediction accuracy.
- However, our experimental results reveal that the determination as to whether items are relevant is not always independent from a sensitive feature.

These results lead to the conclusion that we must develop a new notion of recommendation independence fitting for a find-good-items task.

This paper is organized as follows. In section 2, we formalize the concept of recommendation independence and an IERS task. We show our new method for enhancing recommendation independence in section 3. Our experimental results are shown in section 4. Related work is discussed in section 5, and section 6 concludes our paper.

2 RECOMMENDATION INDEPENDENCE

This section describes a formal definition of recommendation independence and an independence-enhanced recommendation task.

2.1 Definition

To formalize recommendation independence, we need to specify a *sensitive feature*, using the terminology from studies in the fairness-aware data mining literature [5, 17]. We can then attempt to maintain recommendation independence from this sensitive feature, denoted by S. In Sweeny's example of advertisement placement described in section 1, racial information corresponds to a sensitive feature. R represents a recommendation result, which is the degree of relevance to a user's preference used for sorting candidate items in this paper. Based on information theory, the statement "information about a sensitive feature is excluded from the prediction process of the recommendation" describes the condition in which mutual information between R and S is zero. This condition is equivalent to statistical independence between R and S, i.e., $\Pr[R] = \Pr[R|S]$.

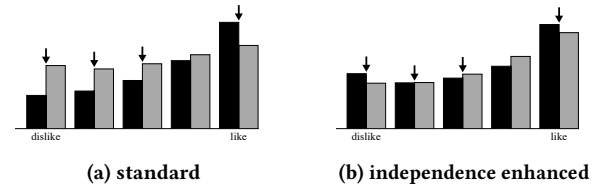(a) standard   (b) independence enhanced

Figure 1: Distributions of the predicted preference scores for each sensitive value

To illustrate the effect of enhancing recommendation independence, we show distributions of predicted preference scores in Figure 1. The charts in this figure show experimental results for ML1M-Year data using an independence parameter, $\eta=10$. The details of the experimental conditions will be shown in section 4. Black and gray bars show the distributions of predicted scores for older and newer movies, respectively. In Figure 1(a), scores are predicted by a standard recommendation algorithm, and older movies are highly rated (see the big gaps between two bars indicated by arrowheads). When recommendation independence is enhanced as in Figure 1(b), the distributions of scores for older and newer movies become much closer (the large gaps are lessened); that is to say, the predicted ratings are less affected by a sensitive feature.

We here note why a sensitive feature must be specified in the definition of recommendation independence. In brief, a sensitive feature must be selected because it is intrinsically impossible to personalize recommendation results if the results are independent of all features. This is due to the *ugly duckling theorem*, which asserts the impossibility of classification without weighing certain features as more important than others [22]. Because recommendation is considered as a task for classifying whether or not items are preferred, certain features inevitably must be weighed. Consequently, it is impossible to enhance independence from all features equally. In the RecSys2011 panel [18], a panelist also pointed out that no information is neutral, and thus individuals are always influenced by information biased in some sense.

2.2 Task Formalization

We formalize a recommendation task whose independence is enhanced. We previously targeted a predicting-ratings recommendation task, which predicted a ratings of items given by a user [4]. In this paper, we concentrate on a find-good-items task, whose goal is to find some items that a user would prefer. $X \in \{1, \ldots, n\}$ and $Y \in \{1, \ldots, m\}$ denote random variables for the user and item, respectively. x and y are instances of X and Y, respectively. We here assume that users explicitly show their preference for items. In a predicting-ratings case, R denotes a random variable that expresses the rating of an item. To fit our previous algorithms for use with a find-good-items task, we make R denote whether an item is relevant or irrelevant to a user. When presenting an item x to a user y, $R=1$ if the item is relevant to the user; otherwise $R=0$. To complete an IERS task, we additionally need a sensitive feature, S, from which independence will be enhanced. The domain of S is currently restricted to a binary type, $\{0, 1\}$, for simplicity.

One training datum consists of a user, x, an item, y, a sensitive value, s (an instance of S), and relevance information, r (an instance of R). A training dataset is the set of N data, $\mathcal{D} =$

$\{(x_i, y_i, s_i, r_i)\}$, $i = 1, \ldots, N$. We define $\mathcal{D}^{(s)}$ as a subset consisting of all data in $\mathcal{D}$ whose sensitive value is s. Given a new datum, (x, y, s), a preference function, $\hat{r}(x, y, s)$, predicts a preference score of the item y for the user x. The aim of an IERS task is to learn this preference function to predict a preference score, indicating the degree of relevance, from a given training dataset under the constraint of recommendation independence. The prediction accuracy generally decreases when an independence constraint is satisfied, due to the loss of usable information. Therefore, it is desirable to satisfy the constraint while sacrificing as little accuracy possible as possible.

3 AN IERS FOR A FIND-GOOD-ITEMS TASK

This section shows a logistic probabilistic-matrix-factorization model. We then introduce an independence-enhanced variant of this model by using a technique in [11].

3.1 A Logistic Matrix Factorization Model

We first introduce a logistic matrix factorization model for a find-good-items task. In our previous algorithms for a predicting-ratings task, we used a probabilistic matrix factorization (PMF) model [14]. Unlike the predicting-ratings case, a target preference, R, can take a value of only 0 or 1 in a find-good-items case. We hence apply a sigmoid function, which is a technique used in [19], and obtain a preference function:

$$\hat{r}(x, y) = \mathrm{sig}\big(\mu + b_x + c_y + \mathbf{p}_x^\top \mathbf{q}_y\big), \qquad (1)$$

where μ, 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. $\mathrm{sig}(a)$ denotes a sigmoid function, $1/(1 + \exp(-a))$. We call this a logistic probabilistic matrix factorization (logistic PMF) model.

3.2 An Independence-Enhanced Logistic PMF Model

We then show an independence-enhanced variant of a logistic PMF model. We use a regularization approach, which was originally developed for a fairness-aware classification task [10]. In this approach, we add an independence term to impose a constraint of recommendation independence. We advocated a simple independence term that was designed to match two means of predicted ratings for $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$ [11].

We first modified a logistic PMF model (1) so that it depended on a sensitive value. For each value of $s \in \{0, 1\}$, we prepared parameter sets, $\mu^{(s)}$, $b_x^{(s)}$, $c_y^{(s)}$, $\mathbf{p}_x^{(s)}$, and $\mathbf{q}_y^{(s)}$. One of the parameter sets was chosen according to the sensitive value, and we obtained the preference function, as follows:

$$\hat{r}(x, y, s) = \mathrm{sig}\Big(\mu^{(s)} + b_x^{(s)} + c_y^{(s)} + \mathbf{p}_x^{(s)\top} \mathbf{q}_y^{(s)}\Big). \qquad (2)$$

We fit this model so as to minimize the following cross-entropy loss, instead of a squared loss used in a predicting-ratings case, because

a domain of R is restricted to 0 or 1:

$$\mathrm{loss}(\mathcal{D}) = \sum_{(x_i, y_i, r_i, s_i) \in \mathcal{D}} -\Big(r_i \log \hat{r}(x_i, y_i, s_i) +$$
$$(1 - r_i) \log(1 - \hat{r}(x_i, y_i, s_i))\Big). \quad (3)$$

Next, we introduce an independence term to impose recommendation independence. This term quantifies the expected degree of independence between a predicted preference and a sensitive feature, with larger values indicating higher levels of independence. The independence term proposed in [11] was designed so as to make the two distributions $\Pr[R|S{=}0]$ and $\Pr[R|S{=}1]$ similar, because R and S become statistically independent if $\Pr[R|S{=}0] = \Pr[R|S{=}1]$. We thus used a squared norm between the means of these distributions, and the independence term became

$$\mathrm{indep}(R, S) = -\left(\frac{\mathbb{S}^{(0)}}{|\mathcal{D}^{(0)}|} - \frac{\mathbb{S}^{(1)}}{|\mathcal{D}^{(1)}|}\right)^2, \qquad (4)$$

where $\mathbb{S}^{(s)}$ is the sum of predicted preferences over the set $\mathcal{D}^{(s)}$,

$$\mathbb{S}^{(s)} = \textstyle\sum_{(x_i, y_i, s_i) \in \mathcal{D}^{(s)}} \hat{r}(x_i, y_i, s_i). \qquad (5)$$

Finally, we defined an objective function used in the regularization approach. The objective function is the sum of a loss term (3), an independence term (4), and an L_2 regularizer:

$$\mathrm{loss}(\mathcal{D}) - \eta \, \mathrm{indep}(R, S) + \lambda \, \mathrm{reg}(\boldsymbol{\Theta}), \qquad (6)$$

where $\eta > 0$ is an independence parameter to balance the loss and independence, $\lambda > 0$ is a regularization parameter, and $\mathrm{reg}(\boldsymbol{\Theta})$ is an L_2 regularizer to avoid over-fitting. By minimizing this objective, the parameters of models can be estimated so that the learned prediction function makes accurate predictions and satisfies the constraint of recommendation independence. Once the parameters of a model are estimated, preference scores for new data can be predicted by a prediction function (2).

4 EXPERIMENTS

We implemented the algorithm in section 3 and applied it to benchmark datasets to inspect the changes in accuracy and independence. Below, we present the details of the datasets and experimental conditions, and then provide experimental results.

4.1 Datasets

We used a Movielens 1M dataset (ML1M) [6] in our experiments. The number of users, items, and ratings were $6,040$, $3,706$, and $1,000,209$, respectively. We regarded a user as preferring an item if the user gave the item a rating of 4 or higher.

We tested two types of sensitive features. The first, Year, represented whether a movie's release year was later than 1990. We selected this feature because it has been proven to influence preference patterns [15]. The sizes of ML1M-Year datasets whose sensitive values were 0 and 1 were $456,683$ and $543,526$, respectively. The second feature, Gender, represented the user's gender. The movie rating depended on the user's gender, and our recommender increased the independence of this information. The sizes of ML1M-Gender datasets whose sensitive values were 0 and 1 were $753,769$ and $246,440$, respectively. Comparing these two sensitive features, the sizes of ML1M-Year datasets divided by sensitive values were

more balanced than those of ML1M-Gender divided by sensitive values. The difference of original mean ratings between datasets, $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$, is about five times larger in the ML1M-Year dataset than in the ML1M-Gender dataset.

4.2 Evaluation Indexes and Experimental Conditions

Next, we evaluated our experimental results in terms of prediction accuracy and the degree of independence. Prediction accuracy was measured by the area under the ROC curve (AUC) [4, 8]. This index measures how much more highly the relevant items are ranked in a recommendation list. A larger value of this index indicates better prediction accuracy.

We adopted two types of independence indexes. The first index measures the degree of independence between a sensitive feature and a preference score derived by equation (2). To evaluate the degree of independence, we checked the equality of the distributions of predicted ratings. For this purpose, we adopted the statistic of the two-sample Kolmogorov-Smirnov test (KS), which is a nonparametric test for the equality of two distributions. The KS statistic is defined as the area between two empirical cumulative distributions of predicted preferences for $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$. A smaller KS indicates that R and S are more independent.

The second type of independence indexes is designed to evaluate the independence of a ranked list. We first assume that candidate items whose predicted preference scores are larger than a threshold are relevant items and the remaining items are irrelevant. A random variable, $\tilde{R}$, represents whether an item is relevant ($\tilde{R} = 1$) or irrelevant ($\tilde{R} = 0$), and $\tilde{r}$ denotes its instance. The degree of independence between two binary variables, S and $\tilde{R}$, was evaluated by the following two indexes. Mutual information (MI) is defined as:

$$\mathrm{MI} = \sum_{\tilde{r} \in \{0,1\}} \sum_{s \in \{0,1\}} \Pr[\tilde{r}, s](\log \Pr[\tilde{r}, s] - \log \Pr[\tilde{r}] \Pr[s]), \quad (7)$$

and becomes 0 if $\tilde{R}$ and S are perfectly independent. Calders & Verwer's discrimination score (CVS) [1] is defined as the probability of being relevant given $S{=}0$ subtracted by that given $S{=}1$,

$$\mathrm{CVS} = \Pr[\tilde{R}{=}1|S{=}1] - \Pr[\tilde{R}{=}1|S{=}0], \quad (8)$$

and becomes 0 if $\tilde{R}$ and S are perfectly independent.

The standard logistic PMF model and independence-enhanced logistic PMF model in section 3 were applied to the datasets in section 4.1. We tuned the hyper-parameters of the model so as to optimize the AUC obtained by a standard logistic PMF model. We used a regularization parameter, $\lambda = 0.1$, and dimension of cross terms, $K = 5$. We changed an independence parameter, η, from 10^{-2} to 10^{2} and observed the accuracy and independence indexes. We performed a five-fold cross-validation procedure to obtain evaluation indexes for the accuracy and independence.

4.3 Experimental Results

In this experiment, we attempted to answer two questions. First, we examined whether or not our method as described in section 3 could actually enhance recommendation independence between a preference score and a sensitive feature. Second, in the case that independence of a preference score was enhanced, we analyzed whether the relevance of items was also independent.

To focus on the first question, whether our independence-enhancement method could enhance recommendation independence, we computed AUC and KS indexes by changing an independence parameter, η. Additionally, we showed the means of predicted preferences for two datasets, $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$, in order to visualize how two the distributions were matched. Figures 2 and 3 show the experimental results. In terms of accuracy, Figures 2(a) and 3(a) show that the loss in accuracy measured by the AUC was very slight. These results were highly contrasted with those of our past experiments, in which the increase rate of error for the predicting-rating task was much higher. This may have been because, although the absolute values of predicted preference scores were changed, the relative rankings of scores among items were preserved. To examine this hypothesis, we compared pairs of predicted scores derived by our algorithms whose independence parameters were $\eta = 0.01$ and $\eta = 10$. The means of absolute differences were 0.053 (Year) and 0.025 (Gender), clearly indicating that the predicted scores were changed. Rank correlations (Spearman's ρ) between pairs of scores were extremely high, 0.978 (Year) and 0.990 (Gender). This observation means that the relative rankings among predicted scores were almost completely preserved, even if recommendation independence was enhanced, and thus the AUCs were not decreased because an AUC index was invariant for any monotonic transformations.

On the other hand, the independence between a predicted preference score and a sensitive feature was clearly enhanced in Figure 2(b). This claim could also be confirmed by the observation that the means of scores derived from $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$ were made increasingly equal by increasing the parameter η in Figure 2(c). In Figure 3(b), it was unclear whether or not the index decreased, because the KS statistics were initially small. However, the matching of the two means in Figure 3(c) proved that the independence was enhanced. From the above, it may be concluded that recommendation independence of a preference score could be enhanced by our logistic PMF model, while the loss in accuracy was very slight.

We were thus able to confirm that the independence of a preference score, R, was enhanced. Next, we moved on to the second question, concerning the independence of the relevance of items from a sensitive feature. As described in section 4.2, we predicted preference scores for all user-item pairs in a dataset in a 5-fold cross-validation procedure, then ranked these items according as their scores are in descending order. In a find-good-items case, the top-k ranked items were assumed to be relevant, and were displayed to users. Hence, we have to take into account the enhancement of independence between a sensitive feature and an event whether a recommended item was relevant ($\tilde{R}{=}1$) or irrelevant ($\tilde{R}{=}0$). We then examine whether or not the enhancement between R and S could enhance the independence between $\tilde{R}$ and S. To examine the independence, we computed the independence indexes as shown in equations (7) and (8) at various threshold of k. Figures 4 and 5 show the changes in the independence indexes according to the number of relevant items, k. By enhancing the independence of preference scores, the independence in regard to relevance was also enhanced for most of the values of k, when compared with a standard recommender. However, the independence of relevance

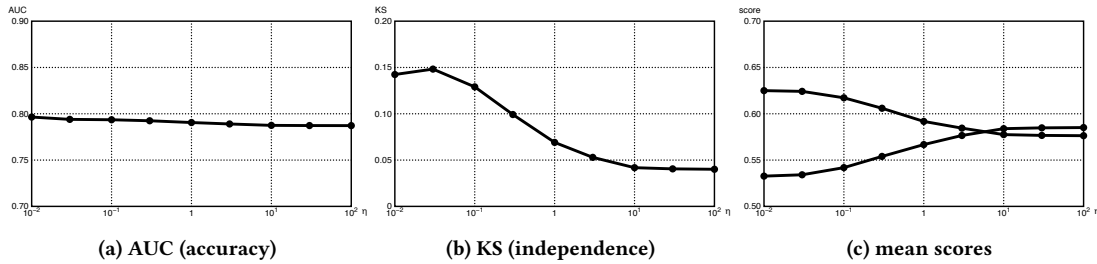| (a) AUC (accuracy) | (b) KS (independence) | (c) mean scores |

Figure 2: Changes of accuracy and independence indexes for the ML1M-Year dataset

NOTE : These figures show the changes of indexes according to an independence parameter, η. The X-axes represent the independence parameter in a logarithmic scale. The Y-axis of the subfigure (a) shows an AUC index to evaluate prediction accuracy. The Y-axis of the subfigure (b) shows the Kolmogorov-Smirnov (KS) statistic to evaluate recommendation independence. Larger AUC indicates better performance in accuracy, and smaller KS indicates better performance in independence. Subfigure (c) shows the means of predicted preference scores for the datasets, $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$.
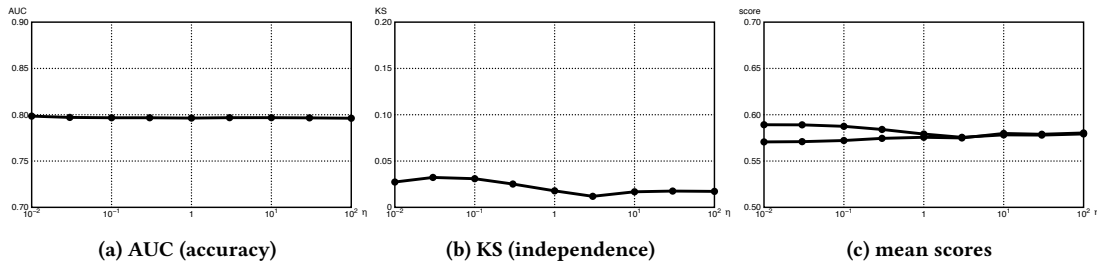

| (a) AUC (accuracy) | (b) KS (independence) | (c) mean scores |

Figure 3: Changes of accuracy and independence indexes for the ML1M-Gender dataset

NOTE : See the note for Figure 2.

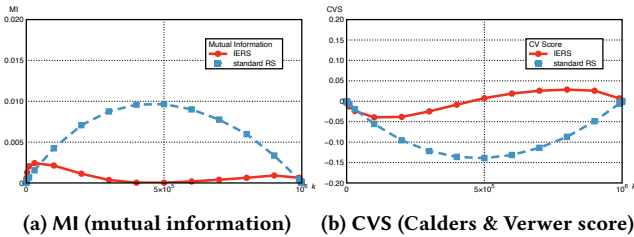| (a) MI (mutual information) | (b) CVS (Calders & Verwer score) |

Figure 4: Changes of independence between $\tilde{R}$ and S for the ML1M-Year dataset

NOTE : These figures show the changes of independence indexes according to the number of relevant items. The X-axes represent the number of relevant items, k. The Y-axis of the subfigure (a) shows mutual information (equation (7)). Blue broken lines show the changes of independence obtained by a standard recommendation algorithm, and red solid lines show the changes obtained by our independence-enhanced recommendation algorithm. A relevance variable, $\tilde{R}$, and a sensitive feature, S, are completely independent if the mutual information is zero. The Y-axis of the subfigure (b) shows Calders and Verwer's discrimination indexes (equation (8)). These indexes are exactly zero if $\tilde{R}$ and S are independent.

was not enhanced for small k in both datasets and indexes. Unfortunately, because users cannot check many items, independence for small k is very important. Therefore, this failure to enhance independence was a serious issue. From this experiment, the enhancement of independence in regard to preference scores did not always enhance independence of relevance.
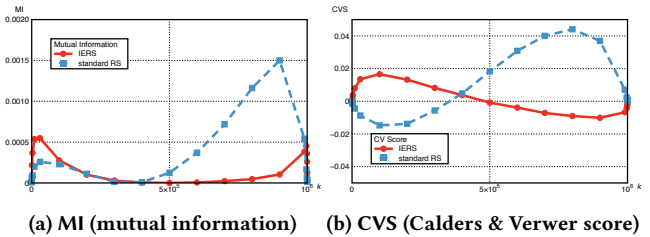

| (a) MI (mutual information) | (b) CVS (Calders & Verwer score) |

Figure 5: Changes of independence between $\tilde{R}$ and S for the ML1M-Gender dataset

NOTE : The note for Figure 4 applies, except that the scaling of Y-axes is changed to clarify the differences of independence indexes.

The experimental results could be summarized as follows:

- Our algorithm could successfully enhance independence between a preference score and a sensitive feature, without appreciably decreasing the accuracy compared to a predicting-ratings case.
- The independence in terms of relevance might not always be enhanced by enhancing the independence of a preference score.

From these experimental results, we conclude that a method must be specially designed to enhance independence between item relevance and sensitive information.

5 RELATED WORK

We wish to emphasize that recommendation independence is distinct from recommendation diversity [16, 23]. First, while diversity may be the property of a set of recommendations, independence is a relation between each recommendation and a sensitive feature. Second, recommendation independence depends on the specification of a sensitive feature, while recommendation diversity depends on the specification of a similarity metric between a pair of items. Finally, while diversity seeks to provide a wider range of topics, independence seeks to provide unbiased information.

We adopted techniques for fairness-aware data mining to enhance the independence. Fairness-aware data mining is a general term for mining techniques designed so that sensitive information does not influence the mining results. Pedreschi et al. first advocated such mining techniques, which emphasized the unfairness in association rules whose consequents include serious determinations [17]. Another technique of fairness-aware data mining focuses on predictions designed so that the influence of sensitive information on the predictions is reduced [1, 10]. These techniques would be directly useful in the development of an independence-enhanced variant of content-based recommender systems, because content-based recommenders can be implemented by standard classifiers. Specifically, class labels indicate whether or not a user prefers an item, and the features of objects correspond to features of the item.

The concept behind recommendation transparency is that it might be advantageous to explain the reasoning underlying individual recommendations. Indeed, such transparency has been proven to improve the satisfaction of users [20], and different methods of explanation have been investigated [7]. In the case of recommendation transparency, the system tries to persuade users of its objectivity by demonstrating that the recommendations were not made by any malicious manipulations. On the other hand, in the case of independence, the objectivity is guaranteed by satisfying a previously defined regulation, i.e., recommendation independence.

6 CONCLUSIONS

We previously developed a method to enhance recommendation independence for a predicting-ratings task. In this paper, we examined recommendation independence for a find-good-items task. We designed a new model to enhance independence of a predicted preference score from a sensitive feature. We empirically showed that this model could enhance independence from a preference score, but the losses in accuracy were very slight. We further examined independence in terms of the relevance of recommended items, but this type of independence sometimes failed to be enhanced.

There are many functionalities required for an IERS. From our experimental results, we must consider a new notion of recommendation independence in terms of a ranked recommendation list for a find-good-items task. Because in this paper we assumed that users explicitly rate the relevance of items, we have to develop a method applicable to the case of implicit ratings. However, it would be difficult to select which items should be treated as irrelevant, because such selection would influence the state of independence. Bayesian extension would not be straightforward because the parameters are probabilistically generated and recommendation independence might be violated under specific choices of parameters. Because

sensitive features are currently restricted to binary types, we will try to deal with sensitive features whose types are multivariate discrete or continuous.

7 ACKNOWLEDGMENTS

REFERENCES

[1] T. Calders and S. Verwer. 2010. Three naive Bayes Approaches for Discrimination-free Classification. *Data Mining and Knowledge Discovery* 21 (2010), 277–292.
[2] Ò. Celma and P. Cano. 2008. 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*.
[3] S. Forden. 2012. Google Said to Face Ultimatum From FTC in Antitrust Talks. Bloomberg. (Nov. 13 2012). ⟨http://bloom.bg/PPNEaS⟩.
[4] A. Gunawardana and G. Shani. 2009. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research* 10 (2009), 2935–2962.
[5] S. Hajian, F. Bonchi, and C. Castillo. 2016. Algorithmic Bias: from Discrimination Discovery to Fairness-Aware Data Mining. The 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, Tutorial. (2016).
[6] F. M. Harper and J. A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. on Interactive Intelligent Systems* 5, 4 (2015).
[7] J. L. Herlocker, J. A. Konstan, and J. Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proc. of the Conf. on Computer Supported Cooperative Work*. 241–250.
[8] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. on Information Systems* 22, 1 (2004), 5–53.
[9] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma. 2012. Enhancement of the Neutrality in Recommendation. In *The 2nd Workshop on Human Decision Making in Recommender Systems*.
[10] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma. 2012. Fairness-aware Classifier with Prejudice Remover Regularizer. In *Proc. of the ECML PKDD 2012, Part II*. 35–50. [LNCS 7524].
[11] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma. 2013. Efficiency Improvement of Neutrality-enhanced Recommendation. In *The 3rd Workshop on Human Decision Making in Recommender Systems*.
[12] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma. 2014. Correcting Popularity Bias by Enhancing Recommendation Neutrality. In *The 8th ACM Conference on Recommender Systems, Poster*.
[13] T. Kamishima, S. Akaho, H. Asoh, and I. Sato. 2016. Model-Based Approaches for Independence-Enhanced Recommendation. In *Proc. of the IEEE 16th Int'l Conf. on Data Mining Workshops*. 860–867.
[14] Y. Koren. 2008. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. In *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. 426–434.
[15] Y. Koren. 2009. Collaborative Filtering with Temporal Dynamics. In *Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. 447–455.
[16] S. M. McNee, J. Riedl, and J. A. Konstan. 2006. Accurate Is Not Always Good: How Accuracy Metrics Have Hurt Recommender Systems. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 1097–1101.
[17] D. Pedreschi, S. Ruggieri, and F. Turini. 2008. Discrimination-aware Data Mining. In *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. 560–568.
[18] P. Resnick, J. Konstan, and A. Jameson. 2011. Panel on The Filter Bubble. The 5th ACM Conf. on Recommender Systems. (2011). ⟨http://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/⟩.
[19] R. Salakhutdinov and A. Mnih. 2008. Probabilistic Matrix Factorization. In *Advances in Neural Information Processing Systems 20*. 1257–1264.
[20] R. Sinha and K. Swearingen. 2002. The Role of Transparency in Recommender Systems. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. 830–831.
[21] L. Sweeney. 2013. Discrimination in Online Ad Delivery. *Commun. ACM* 56, 5 (2013), 44–54.
[22] S. Watanabe. 1969. *Knowing and Guessing – Quantitative Study of Inference and Information*. John Wiley & Sons.
[23] C. N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *Proc. of the 14th Int'l Conf. on World Wide Web*. 22–32.