

Model-Based Approaches for Independence-Enhanced Recommendation

Toshihiro Kamishima*, Shotaro Akaho*, Hideki Asoh*, and Issei Sato†

*National Institute of Advanced Industrial Science and Technology (AIST),

AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki, 305-8568 Japan,

Email: mail@kamishima.net (<http://www.kamishima.net/>), s.akaho@aist.go.jp, and h.asoh@aist.go.jp

†The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan, Email: sato@k.u-tokyo.ac.jp

Abstract—This paper studies a new approach to enhance recommendation independence. Such approaches are useful in ensuring adherence to laws and regulations, fair treatment of content providers, and exclusion of unwanted information. For example, recommendations that match an employer with a job applicant should not be based on socially sensitive information, such as gender or race, from the perspective of social fairness. An algorithm that could exclude the influence of such sensitive information would be useful in this case. We previously gave a formal definition of recommendation independence and proposed a method adopting a regularizer that imposes such an independence constraint. As no other options than this regularization approach have been put forward, we here propose a new model-based approach, which is based on a generative model that satisfies the constraint of recommendation independence. We apply this approach to a latent class model and empirically show that the model-based approach can enhance recommendation independence. Recommendation algorithms based on generative models, such as topic models, are important, because they have a flexible functionality that enables them to incorporate a wide variety of information types. Our new model-based approach will broaden the applications of independence-enhanced recommendation by integrating the functionality of generative models.

Keywords—fairness-aware data mining, recommender system, topic model

I. INTRODUCTION

A recommender system searches for items or information predicted to be useful to a user based on the user's previous behaviors and other related information. Recommender systems have become increasingly popular, and their influence is growing. Recommender systems and other personalization technologies have become indispensable tools in support of decision-making.

To avoid unfairness or bias in decision-making due to recommender systems, the influence of specific information should be excluded from the prediction process of recommendation. In other words, independence between specific information and recommendation results 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 [1]. 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 [2]. In this case, no deliberate manipulation was found. However, algorithms that can explicitly exclude information (for example, 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 [3], is a well-known drawback of recommenders. If information on 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 techniques to exclude the influence of specific information, we formalized the notion of recommendation independence and developed algorithms to enhance it. For this purpose, we exploited a technique developed for fairness-aware data mining [4], [5], whose goal is to analyze data while taking into account potential issues of fairness. *Recommendation independence* was formally defined as statistical independence between a recommendation result and specified information. Based on this definition, we developed an *Independence-Enhanced Recommender System* (IERS) that could maintain recommendation independence [6], [7]. In this system, we proposed a regularization approach, in which a regularizer was adopted that imposes a constraint of independence when training a recommendation model.

No approach other than this regularization approach is currently available for enhancing recommendation independence. Although recommendation independence has been successfully enhanced by a regularization approach, it cannot be said that the approach is effective or efficient in every situation where recommender systems are currently used. Designers of recommender systems must consider multiple objectives in addition to accuracy, such as diversity, privacy, and transparency. A regularization approach might

not be simultaneously applicable to the model designed to achieve these objectives. Hence, providing another option for independence enhancement will help in designing more helpful recommender systems.

We hence propose a new model-based approach to enhancing recommendation independence. In this approach, a sensitive variable is added to a generative model for recommendation, such as a topic model, so that it satisfies the constraint of recommendation independence. Many types of recommendation algorithms with generative models have been developed to realize such as the following functionalities. Collaborative filtering and content-based methods are hybridized [8]. Relative preference data collected by the pairwise comparison method can be processed [9]. Social information between users is exploited for better recommendation [10]. Transparency of recommendation is increased by assigning badges to users [11]. The recommendation independence of such models can be enhanced by our new approach, and this should contribute to improving users' satisfaction.

We applied this model-based approach to a recommendation algorithm with a generative model, and examined the characteristics of this new approach by comparing it with our previous regularization approach. In this paper, we extended a latent class model [12], and embedded a sensitive variable while maintaining its independence from an item-rating variable. Because the sensitive variables were embedded in two different ways, two types of independence-enhanced models were developed. We tested these two new models on several benchmark datasets, and compared the behaviors of them with those with our previous method.

Our contributions can be summarized as follows.

- We propose a new model-based approach for enhancing recommendation independence. This provides yet another option when designing an IERS.
- We applied this new approach to a latent class model, and developed two types of models.
- We examined the characteristics of these model-based approaches by comparing their behaviors with those of our previous regularization approach.

This paper is organized as follows. In section II, we formalize the concept of recommendation independence and an IERS task. We show our previous regularization approach and new model-based approach in section III. Experimental results are shown in section IV. Section V covers related work, and section VI concludes our paper.

II. RECOMMENDATION INDEPENDENCE

In this section, we provide a formal definition of *recommendation independence*, and then define a task for an *independence-enhanced recommender system* (IERS).

A. Definition

To formalize recommendation independence, we need to specify a *sensitive feature*, using the terminology in the fairness-aware data mining literature. We then attempt to maintain recommendation independence in relation to this sensitive feature, denoted by S . In Sweeny's example of advertisement placement described in section I, racial information corresponds to a sensitive feature. R represents a recommendation result, which is typically a rating score or an indicator of whether or not a specified item is preferred. 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 the mutual information between R and S is zero. This means that we know nothing about R if we obtain information about S , because information about S is excluded from R . This condition is equivalent to statistical independence between R and S , i.e., $\Pr[R] = \Pr[R|S]$.

B. Task Formalization

We here formalize a recommendation task whose independence is enhanced. We concentrate on a predicting-ratings task [13] and formalize an independence-enhanced variant of the task. $X \in \{1, \dots, n\}$ and $Y \in \{1, \dots, m\}$ denote random variables for the user and item, respectively. An event (x, y) is an instance of a pair (X, Y) . R denotes a random variable for a rating of Y given by X , and r is its instance. These variables are the same as those in an original predicting-ratings task. In the case of an IERS task, we additionally introduce a random sensitive variable, S , which indicates the sensitive feature with respect to which the independence is enhanced, while s denotes an instance of S , with the domain of S restricted to a binary type, $\{0, 1\}$, for simplicity.

One training datum consists of an event, (x, y) , a sensitive value for that event, s , and a rating value for the event, r . A training dataset is the set of N data, $\mathcal{D} = \{(x_i, y_i, s_i, r_i)\}$, $i = 1, \dots, N$. We define $\mathcal{D}^{(s)}$ as a subset that consists of all data in \mathcal{D} whose sensitive value is s . Given a new event, (x, y) , and its corresponding sensitive value, s , a rating prediction function, $\hat{r}(x, y, s)$, predicts a rating of the item y by the user x . The aim of an IERS task is to learn this rating prediction function 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.

III. TWO APPROACHES FOR ENHANCING RECOMMENDATION INDEPENDENCE

This section shows two approaches to independence enhancement. The first is our previous approach [7] of adopting a regularizer that imposes an independence constraint. The

second is our new model-based approach that uses graphical models that satisfy an independence constraint.

A. A Regularization Approach

We here introduce an earlier implementation of an IERS, described in a previous paper [7]. Fundamentally, this approach is the same as that developed for a fairness-aware classification task [14]. However, because the target of the prediction is a continuous rating instead of a discrete class, the application is not straightforward. This approach is to optimize an objective function composed of loss and independence terms. The loss term represents how a model doesn't fit a given training dataset and is minimized to acquire a model that can make more accurate predictions. In our previous work, we adopted a probabilistic matrix factorization (PMF) model [15], [16], and an objective function of the PMF model was used as this loss term. The independence term is used as a regularizer to impose recommendation independence. We advocated a simple independence term that was designed to make two means of predicted ratings agree, each corresponding to a distinct sensitive value.

We show the loss term of this objective function. We adopted a PMF model [15], [16] to predict ratings:

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

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. We modified this model so that it depended on the sensitive value, s . For each value of s , 0 and 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 rating prediction function, as follows:

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

We then adopted squared loss with an L_2 regularizer:

$$\text{loss}(\mathcal{D}) = \sum_{(x_i, y_i, r_i, s_i) \in \mathcal{D}} (r_i - \hat{r}(x_i, y_i, s_i))^2 + \lambda \text{reg}(\Theta), \quad (3)$$

where $\lambda > 0$ is a regularization parameter, and $\text{reg}(\Theta)$ is an L_2 regularizer to avoid over-fitting.

Next, we introduced the other term, the independence term. The independence term quantified the expected degree of independence between the predicted ratings and sensitive values, with larger values indicating higher levels of independence. The independence term proposed in [7] 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

$$\text{indep}(R, S) = -\left(\frac{\mathbb{S}^{(0)}}{N^{(0)}} - \frac{\mathbb{S}^{(1)}}{N^{(1)}}\right)^2, \quad (4)$$

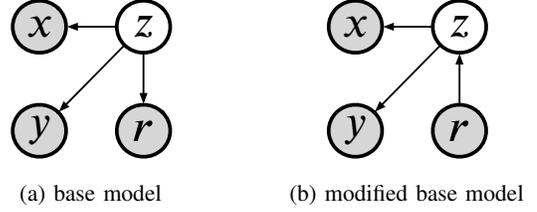


Figure 1: Base latent class models

where $N^{(s)} = |\mathcal{D}^{(s)}|$ and $\mathbb{S}^{(s)}$ is the sum of predicted ratings over the set $\mathcal{D}^{(s)}$,

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

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

$$\text{loss}(\mathcal{D}) - \eta \text{indep}(R, S), \quad (6)$$

where $\eta > 0$ is an independence parameter to balance the loss and independence. 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, ratings for new data can be predicted by a prediction function (2).

B. A Model-Based Approach

We now propose our new approach, a model-based approach, which imposes recommendation independence using a generative model in which a rating variable is independent from a sensitive variable. Because our new model is an extension of a latent class model [12], we introduce this model first. We extend this base model by embedding a sensitive variable while maintaining independence between the sensitive variable and the rating variable.

We now introduce a latent class model (LCM) for the collaborative filtering proposed by Hofmann [12]. This is a variant of probabilistic latent semantic analysis [17], so it can capture ratings. While Hoffmann showed several types of latent class models, we focused on the type (b) model in Figure 1 of the original article. Graphical representation of this model is shown here in Figure 1(a), and its corresponding joint distribution is

$$\Pr[x, y, r] = \sum_z \Pr[x|z] \Pr[y|z] \Pr[r|z] \Pr[z]. \quad (7)$$

Z is a latent random variable that represents types of preference patterns, and z stands for an instance of Z . The model's assumption that X , Y , and R are mutually independent given Z drastically reduces the number of model parameters. Z is a discrete variable whose domain is $\{1, \dots, K\}$. All distributions in the rhs of equation (7), $\Pr[X|Z]$, $\Pr[Y|Z]$, $\Pr[R|Z]$, and $\Pr[Z]$, follow categorical

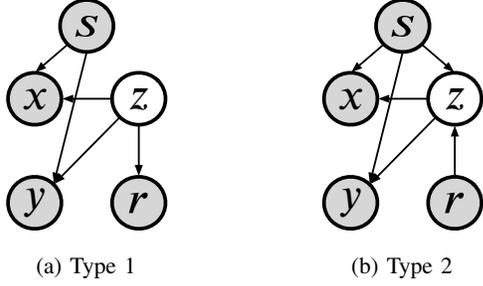


Figure 2: Generative models for independence-enhanced recommendation

distributions. In this model, R is a discrete variable that can take $|R|$ levels of ratings.

Below, we develop two types of generative models for independence-enhanced recommendation.

1) *Type 1 Model*: Figure 2(a) depicts the type 1 model. A sensitive variable, S , is embedded while maintaining the independence between R and S . In addition, the model assumption of the base model, that is the mutual independence between X , Y , and R given Z , is also maintained to reduce the number of model parameters. To let the preference patterns of users for items be influenced by a sensitive feature, we designed the model so that X and Y depend on S . Note that we did not make R and Z dependent on S , because such dependency violates the independence between R and S . A joint distribution of the type 1 model thus becomes

$$\Pr[x, y, r, s] = \sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[s] \Pr[z]. \quad (8)$$

All factor distributions in the rhs follow categorical distributions.

Parameters of the type 1 model can be estimated by an EM algorithm. An EM algorithm alternately iterates an E-step and an M-step until the estimation of parameters converges. In an E-step, the conditional distribution of a latent variable given observed variables is computed using the current parameters:

$$\Pr[z|x, y, r, s] \leftarrow \frac{\Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[z]}{\sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[z]}. \quad (9)$$

In an M-step, all parameters are updated using this new distribution (9):

$$\begin{aligned} \Pr[x|s, z] &\leftarrow \frac{\sum_D \{I[x=x_i, s=s_i] q_i(z)\} + \alpha}{\sum_D \{I[s=s_i] q_i(z)\} + n\alpha}, \\ \Pr[y|s, z] &\leftarrow \frac{\sum_D \{I[y=y_i, s=s_i] q_i(z)\} + \alpha}{\sum_D \{I[s=s_i] q_i(z)\} + m\alpha}, \\ \Pr[r|z] &\leftarrow \frac{\sum_D \{I[r=r_i] q_i(z)\} + \alpha}{\sum_D \{q_i(z)\} + |R|\alpha}, \\ \Pr[s] &\leftarrow \frac{\sum_D I[s=s_i] + \alpha}{N + 2\alpha}, \text{ and} \\ \Pr[z] &\leftarrow \frac{\sum_D \{q_i(z)\} + \alpha}{N + K\alpha}, \end{aligned} \quad (10)$$

where x_i , y_i , r_i , and s_i are components of the i -th data in \mathcal{D} , and $q_i(z) = \Pr[z|x_i, y_i, r_i, s_i]$. α is a parameter of Laplace smoothing to avoid a zero-counting problem. $I[\cdot]$ is an indicator function that takes 1 if all conditions in arguments are satisfied and otherwise is 0.

Once model parameters are estimated from a training dataset, a rating for a new datum (x, y, s) can be predicted by a rating prediction function

$$\begin{aligned} \hat{r}(x, y, s) &= E_{\Pr[r|x, y, s]}[\text{level}(r)] \\ &= \sum_r \Pr[r|x, y, s] \text{level}(r), \end{aligned} \quad (11)$$

where $\text{level}(r)$ denotes the r -th rating value, and the distribution of R given a new datum is

$$\Pr[r|x, y, s] = \frac{\sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[z]}{\sum_{r, z} \Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[z]}. \quad (12)$$

2) *Type 2 Model*: We now move on to the type 2 model in Figure 2(b). To derive this type 2 model, the original base model is rewritten as in Figure 1(b), and its corresponding joint distribution becomes

$$\Pr[x, y, r] = \sum_z \Pr[x|z] \Pr[y|z] \Pr[z|r] \Pr[r]. \quad (13)$$

In this rewritten model, the assumption of mutual independence between X , Y , and R given Z holds. We again add a sensitive variable while maintaining the independence between R and S , and the assumption of the base model. We get

$$\Pr[x, y, r, s] = \sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[r|z] \Pr[r] \Pr[s] \Pr[z|r, s]. \quad (14)$$

All factor distributions follow the categorical distributions. We expect that a type 2 model can more strictly enhance recommendation independence, because a sensitive feature, S , is directly dependent on Z as well as X and Y .

Parameters of this model can be also estimated by the following EM algorithm.

E-step:

$$\Pr[z|x, y, r, s] \leftarrow \frac{\Pr[x|s, z] \Pr[y|s, z] \Pr[z|r, s]}{\sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[z|r, s]}. \quad (15)$$

M-step:

$$\begin{aligned} \Pr[x|s, z] &\leftarrow \frac{\sum_D \{I[x=x_i, s=s_i] q_i(z)\} + \alpha}{\sum_D \{I[s=s_i] q_i(z)\} + n\alpha}, \\ \Pr[y|s, z] &\leftarrow \frac{\sum_D \{I[y=y_i, s=s_i] q_i(z)\} + \alpha}{\sum_D \{I[s=s_i] q_i(z)\} + m\alpha}, \\ \Pr[r] &\leftarrow \frac{\sum_D I[r=r_i] + \alpha}{N + |R|\alpha}, \\ \Pr[s] &\leftarrow \frac{\sum_D I[s=s_i] + \alpha}{N + 2\alpha}, \text{ and} \\ \Pr[z|r, s] &\leftarrow \frac{\sum_D \{I[r=r_i, s=s_i] q_i(z)\} + \alpha}{\sum_D I[r=r_i, s=s_i] + K\alpha}. \end{aligned} \quad (16)$$

By using estimated parameters, a rating value for a new datum can again be predicted by equation (11) except for the distribution of ratings:

$$\Pr[r|x, y, s] = \frac{\sum_z \Pr[x|s, z] \Pr[y|s, z] \Pr[r] \Pr[z|r, s]}{\sum_{r, z} \Pr[x|s, z] \Pr[y|s, z] \Pr[r] \Pr[z|r, s]}. \quad (17)$$

Table I: Summary of datasets

data	ML1M	Flixster	Sushi
# of users	6,040	147,612	5,000
# of items	3,706	48,794	100
# of ratings	1,000,209	8,196,077	50,000
rating scale	1, 2, ..., 5	.5, 1, ..., 5	0, 1, ..., 4
mean rating	3.58	3.61	1.27

Table II: Sizes, means, and variances of data subsets for each sensitive value

Datasets	size		mean		variance	
	$S=0$	$S=1$	$S=0$	$S=1$	$S=0$	$S=1$
ML1M-Year	456,683	543,526	3.72	3.46	1.15	1.30
ML1M-Gender	753,769	246,440	3.57	3.62	1.25	1.23
Flixster	3,868,842	4,327,235	3.71	3.53	1.18	1.18
Sushi-Age	3,150	46,850	2.39	2.75	1.75	1.60
Sushi-Gender	23,730	26,270	2.80	2.66	1.43	1.77
Sushi-Seafood	43,855	6,145	2.76	2.46	1.61	1.58

IV. EXPERIMENTS

We implemented our independence-enhanced recommender systems and applied them to benchmark datasets. Below, we present the details regarding these datasets and experimental conditions, then compare the performances of two standard models and three independence-enhanced models.

A. Datasets and Evaluation Indexes

We can now consider the experimental conditions in detail, including the datasets, methods, and evaluation indexes. We used the three datasets summarized in Tables I and II. The first dataset was the Movielens 1M dataset (ML1M) [18]. We tested two types of sensitive features for this set. 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 [19]. 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 second set was the larger Flixster dataset¹ [20]. Because neither the user nor the item features were available, we adopted the popularity of items as a sensitive feature. Candidate movies were first sorted by the number of users who rated the movie in a descending order, with the sensitive feature being whether or not a movie was in the top 1% of this list. A total of 47.2% of ratings were assigned to this top 1% of items. The third dataset was our **Sushi** dataset² [21], for which we adopted three types of sensitive features: **Age** (the user was a teen or

¹<http://www.cs.ubc.ca/~jamalim/datasets/>

²<http://www.kamishima.net/sushi/>

older), **Gender** (the user was male or female), and **Seafood** (whether or not a type of sushi was seafood). We selected these features because the means of rating between two subsets, $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$, diverged.

We evaluated our experimental results in terms of prediction errors and the degree of independence. Prediction errors were measured by the mean absolute error (MAE) [13]. This index was defined as the mean of the absolute difference between the observed ratings and predicted ratings. A smaller value of this index indicates better prediction accuracy. To evaluate the degree of independence, we checked the equality of the distributions of predicted ratings. As an independence index, 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 ratings for $\mathcal{D}^{(0)}$ and $\mathcal{D}^{(1)}$. A smaller KS indicates that R and S are more independent.

B. Tested Models and Experimental Conditions

We tested the prediction accuracies and the degree of independence of two standard recommendation models and three independence-enhanced models. As standard methods, whose independence was not enhanced, we chose a PMF model (equation (1)), which was a base model of our previous regularization approach, and a latent class model (equation (7)), which was a base model of our new model-based approach. We refer to these two standard models as **pmf** and **lcm**, respectively. The first independence-enhanced model was our previous model proposed in [7], whose objective function is equation (6). We refer to this model as **pmf-r**. The other two models were our new models: Type 1 (equation (8)) and Type 2 (equation (14)) models. We abbreviate these models as **lcm-mb1** and **lcm-mb2**, respectively. Program Codes are available at <http://www.kamishima.net/iers/>.

The experimental conditions of these five models were as follows. The parameters of **pmf** and **pmf-r** models were optimized by the conjugate gradient method implemented in the SciPy package [22]. These models have two hyper-parameters: the number of latent factors K , and a regularization parameter λ . We used $\{K=7, \lambda=1\}$ for the ML1M datasets, $\{K=20, \lambda=30\}$ for the Flixster dataset, and $\{K=5, \lambda=10\}$ for the Sushi datasets. We let $\eta=100$ be the independence parameter; this implies the independence was greatly enhanced. The hyper-parameters of **lcm**, **lcm-mb1**, and **lcm-mb2** were the number of dimensions of the latent variable, K , and the Laplace smoothing parameter, α . The dimension parameter, K , was 10 for the ML1M and Flixster datasets, and was 5 for the Sushi datasets. As the smoothing parameter, we used $\alpha = 0.1$ for all datasets. Note that these hyper-parameters were chosen so that standard models were well performed in MAE. We performed a five-fold

cross-validation procedure to obtain evaluation indexes of the accuracy and independence indexes.

C. Experimental Results

We here compared two standard and three independence-enhanced models. Table III summarizes accuracies measured by MAE and the degrees of independence measured by KS obtained by five methods on six datasets.

In this experiment, we examine the following questions. First, we checked whether independence-enhanced methods could really enhance the recommendation independence compared with corresponding standard methods. Such enhancement was not theoretically guaranteed because we used optimization methods that could fail to find the global optima. Second, we compared the characteristics of three independence-enhanced models. We specifically inspected the difference between regularization and model-based approaches and the differences between the two model-based approaches.

We first focused on the first question, whether independence-enhancement methods succeed at enhancing recommendation independence. As observed in our previous work [7], pmf-r could greatly reduce a KS index, implying enhancement of independence, except for the Sushi-Seafood case. Note that the poor performance in the Sushi-Seafood case would be due to the small size of the dataset such that $S=1$ as in Table II. On the other hand, the accuracies were not very worsened, and pmf-r could achieve a good trade-off between accuracy and independence. Regarding lcm families, both type 1 and 2 models successfully enhanced recommendation independence, compared with the base lcm model. Furthermore, accuracies were greatly improved or only slightly worsened; this was very preferable property.

We turn to the second question. Characteristics of three independence-enhanced models were examined. Some differences between the regularization and model-based approaches were observed. Generally speaking, lcm variants were worse than pmf-r in accuracy, because a standard lcm was inferior to pmf. Though both of them are basically matrix factorization models, the lcm model is more restricted because parameters must satisfy the constraint of probability. Regarding independence, lcm families and pmf were comparable on the small Sushi dataset, but pmf performed better on the other larger datasets. According to our previous analysis in fairness-aware classification [23], deterministic transformation in class decision affected independence. An operation of expectation in the rating prediction function might influence the performance in independence. Note that it is not straightforward to apply a regularization approach to the lcm model due to the existence of a latent variable, so we have not checked this hypothesis yet.

We then compared the two lcm models. We expected that the type 2 model would perform better for recommendation

independence, because of the direct dependence of S on Z . However, these two methods were almost equivalent in terms of accuracy and independence. We changed the hyperparameters, but the behaviors of these two models remained roughly the same. This would be because the assumptions of these two models, recommendation independence and mutual independence between X , Y , and R given Z , were fundamentally the same, though their model parameterizations were different.

From the above, it may be concluded that recommendation independence could be enhanced by our new model-based approach. The performances of model-based and regularization approaches are comparable for smaller datasets, while the previous approach performed better on larger datasets.

V. RELATED WORK

We briefly survey techniques for promoting diversity, because independence is related to recommendation diversity, which is an attempt to recommend a set of items that are mutually dissimilar. McNee et al. pointed out that recommendation diversity is important, because users become less satisfied with recommended items if similar items are repeatedly shown [24]. To our knowledge, Ziegler et al. were the first to propose diversifying recommendation lists by selecting items less similar to those already selected [25]. Zhang et al. formalized a method for composing a recommendation list as an optimization problem with a penalty term to discourage the selection of similar items [26]. Lathia et al. discussed the concept of temporal diversity that is not defined in a single recommendation list, but over temporally successive recommendations [27]. Adomavicius et al. discussed the aggregate diversity of items that are recommended to a whole population of users [28].

We wish to emphasize that recommendation independence is distinct from recommendation diversity due to the following reasons. 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 independence [5]. 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 [4]. Another technique of fairness-aware data mining focuses on classifications designed so that the influence of sensitive information on the classification results is reduced [14], [29],

Table III: Comparison of accuracies and independence indexes derived by standard and independence-enhanced recommendation methods

method	ML1M				Flixster		Sushi					
	Year		Gender		MAE	KS	Age		Gender		Seafood	
	MAE	KS	MAE	KS			MAE	KS	MAE	KS	MAE	KS
pmf	0.685	0.1687	0.685	0.0389	0.655	0.1523	0.906	0.1114	0.906	0.0496	0.906	0.2756
pmf-r	0.697	0.0271	0.694	0.0050	0.653	0.0165	0.926	0.0256	0.925	0.0321	0.925	0.2002
lcm	0.729	0.1984	0.729	0.0487	0.671	0.1787	0.916	0.1754	0.916	0.0714	0.916	0.1838
lcm-mb1	0.717	0.0752	0.719	0.0243	0.672	0.0656	0.919	0.0335	0.912	0.0350	0.915	0.1953
lcm-mb2	0.720	0.1030	0.720	0.0364	0.672	0.0656	0.919	0.0329	0.912	0.0349	0.915	0.1953

[30], [31]. 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 target user prefers a target item, and the features of objects correspond to features of item contents.

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 [32], and different methods of explanation have been investigated [33]. In the case of recommendation transparency, the system convinces users of its objectivity by demonstrating that the recommendations were made without intentional bias. In the case of independence, the objectivity is guaranteed based on mathematically defined principles.

Because independence-enhanced recommender systems can be used to avoid the exploitation of private information, these techniques are related to privacy-preserving data mining [34]. To protect the private information contained in rating information, dummy ratings are added [35]. Privacy attack strategies have been used as a tool for detecting discrimination [36], [37].

VI. CONCLUSIONS

In our previous work, we defined the notion of recommendation independence and developed a method to enhance it. In this paper, we propose a new model-based approach to provide yet another option for independence enhancement. We applied this approach to a latent class recommendation model. We empirically showed that this new model could successfully enhance recommendation independence.

There are many functionalities required for an IERS. Bayesian extension of a model-based approach would not straightforward because 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 also try to develop independence-enhancement approaches that can deal with a sensitive feature, be it multivariate

discrete or continuous. Although our current technique is mainly applicable to the task of predicting ratings, we plan to develop another algorithm for the task of recommending good items.

ACKNOWLEDGMENT

We would like to thank for providing datasets for the Grouplens research lab and Dr. Mohsen Jamali. This work is supported by MEXT/JSPS KAKENHI Grant Number JP24500194 and JP15K00327, and JP16H02864.

REFERENCES

- [1] L. Sweeney, "Discrimination in online ad delivery," *Communications of the ACM*, vol. 56, no. 5, pp. 44–54, 2013.
- [2] S. Forden, "Google said to face ultimatum from FTC in antitrust talks," Bloomberg, Nov. 13 2012, (<http://bloom.bg/PPNEaS>).
- [3] Ö. 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.
- [4] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-aware data mining," in *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 2008, pp. 560–568.
- [5] S. Hajian, F. Bonchi, and C. Castillo, "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] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "Enhancement of the neutrality in recommendation," in *Proc. of the 2nd Workshop on Human Decision Making in Recommender Systems*, 2012, pp. 8–14.
- [7] —, "Efficiency improvement of neutrality-enhanced recommendation," in *Proc. of the 3rd Workshop on Human Decision Making in Recommender Systems*, 2013, pp. 1–8.
- [8] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *Proc. of the 25th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, 2002, pp. 253–260.

- [9] R. Jin, L. Si, and C.-X. Zhai, "Preference-based graphic models for collaborative filtering," in *Uncertainty in Artificial Intelligence 19*, 2003, pp. 329–336.
- [10] M. Jamali, T. Huang, and M. Ester, "A generalized stochastic block model for recommendation in social rating networks," in *Proc. of the 5th ACM Conf. on Recommender Systems*, 2011, pp. 53–60.
- [11] K. El-Arini, U. Paquet, R. Herbrich, J. V. Gael, and B. Agüera y Arcas, "Transparent user models for personalization," in *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 2012.
- [12] T. Hofmann and J. Puzicha, "Latent class models for collaborative filtering," in *Proc. of the 16th Int'l Joint Conf. on Artificial Intelligence*, 1999, pp. 688–693.
- [13] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *Journal of Machine Learning Research*, vol. 10, pp. 2935–2962, 2009.
- [14] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "Fairness-aware classifier with prejudice remover regularizer," in *Proc. of the ECML PKDD 2012, Part II*, 2012, pp. 35–50, [LNCS 7524].
- [15] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems 20*, 2008, pp. 1257–1264.
- [16] Y. Koren, "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*, 2008, pp. 426–434.
- [17] T. Hofmann, "Probabilistic latent semantic analysis," in *Uncertainty in Artificial Intelligence 15*, 1999, pp. 289–296.
- [18] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Trans. on Interactive Intelligent Systems*, vol. 5, no. 4, 2015.
- [19] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 2009, pp. 447–455.
- [20] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. of the 4th ACM Conf. on Recommender Systems*, 2010, pp. 135–142.
- [21] T. Kamishima, "Nantonac collaborative filtering: Recommendation based on order responses," in *Proc. of The 9th Int'l Conf. on Knowledge Discovery and Data Mining*, 2003, pp. 583–588.
- [22] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2000-, (<http://www.scipy.org/>).
- [23] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "The independence of the fairness-aware classifiers," in *Proc. of the 4th IEEE Int'l Workshop on Privacy Aspects of Data Mining*, 2013, pp. 849–858.
- [24] 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*, 2006, pp. 1097–1101.
- [25] 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*, 2005, pp. 22–32.
- [26] M. Zhang and N. Hurley, "Avoiding monotony: Improving the diversity of recommendation lists," in *Proc. of the 2nd ACM Conf. on Recommender Systems*, 2008, pp. 123–130.
- [27] N. Lathia, S. Hailes, L. Capra, and X. Amatriain, "Temporal diversity in recommender systems," in *Proc. of the 33rd Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, 2010, pp. 210–217.
- [28] G. Adomavicius and Y. Kwon, "Improving aggregate recommendation diversity using ranking-based techniques," *IEEE Trans. on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 896–911, 2012.
- [29] K. Fukuchi, J. Sakuma, and T. Kamishima, "Prediction with model-based neutrality," in *Proc. of the ECML PKDD 2013, Part II*, 2013, pp. 499–514, [LNCS 8189].
- [30] T. Calders and S. Verwer, "Three naive Bayes approaches for discrimination-free classification," *Data Mining and Knowledge Discovery*, vol. 21, pp. 277–292, 2010.
- [31] F. Kamiran, A. Karim, and X. Zhang, "Decision theory for discrimination-aware classification," in *Proc. of the 12th IEEE Int'l Conf. on Data Mining*, 2012, pp. 924–929.
- [32] R. Sinha and K. Swearingen, "The role of transparency in recommender systems," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, 2002, pp. 830–831.
- [33] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proc. of the Conf. on Computer Supported Cooperative Work*, 2000, pp. 241–250.
- [34] C. C. Aggarwal and P. S. Yu, Eds., *Privacy-Preserving Data Mining: Models and Algorithms*. Springer, 2008.
- [35] U. Weinsberg, S. Bhagat, S. Ioannidis, and N. Taft, "Blurme: Inferring and obfuscating user gender based on ratings," in *Proc. of the 6th ACM Conf. on Recommender Systems*, 2012, pp. 195–202.
- [36] S. Hajian, J. Domingo-Ferrer, and O. Farràs, "Generalization-based privacy preservation and discrimination prevention in data publishing and mining," *Data Mining and Knowledge Discovery*, 2014.
- [37] S. Ruggieri, S. Hajian, F. Kamiran, and X. Zhang, "Anti-discrimination analysis using privacy attack strategies," in *Proc. of the ECML PKDD 2014, Part II*, 2014, pp. 694–710, [LNCS 8725].