Recommendation with k-Anonymized Ratings

Jun Sakuma[*,**], **Tatsuya Osame**[*]

[*] University of Tsukuba, 1-1-1 Tennoh-dai, Tsukuba, Ibaraki, 305-8573, Japan.

[**] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan.

E-mail: jun@cs.tsukuba.ac.jp,osame@mdl.cs.tsukuba.ac.jp

Received 25 October 2017; received in revised form 2 December 2017; accepted 4 March 2018

Abstract. Recommender systems are widely used to predict personalized preferences of goods or services using users' past activities, such as item ratings or purchase histories. If collections of such personal activities were made publicly available, they could be used to personalize a diverse range of services, including targeted advertisement or recommendations. However, there would be an accompanying risk of privacy violations. The pioneering work of Narayanan et al. demonstrated that even if the identifiers are eliminated, the public release of user ratings can allow for the identification of users by those who have only a small amount of data on the users' past ratings.

In this paper, we assume the following setting. A collector collects user ratings, then anonymizes and distributes them. A recommender constructs a recommender system based on the anonymized ratings provided by the collector. Based on this setting, we exhaustively list the models of recommender systems that use anonymized ratings. For each model, we then present an item-based collaborative filtering algorithm for making recommendations based on anonymized ratings. Our experimental results show that an item-based collaborative filtering based on anonymized ratings can perform better than collaborative filterings with non-anonymized ratings under certain conditions. This surprising result indicates that, in some settings, privacy protection does not necessarily reduce the usefulness of recommendations. From the experimental analysis of this counterintuitive result, we observed that the sparsity of the ratings could be reduced by anonymization and the variance of the prediction error can be reduced if k, the anonymization parameter, is appropriately tuned.

Keywords. Recommendation, anonymization, privacy, collaborative filtering.

1 Introduction

With rapid advances in online services, a huge amount of information describing users' personal activities is being collected and stored. Recommender systems that use records of user activities are widely used in diverse services, including e-commerce and information dissemination. On the other hand, records describing detailed user activities can sometimes compromise privacy. For example, Narayanan et al. [6] have reported an anonymity assessment of a dataset containing user ratings of movies, which was published as part of competition for recommendation algorithms. They demonstrated that even if the identifiers (i.e., the names of the users) are eliminated from the dataset, the public release of the dataset can allow for the reidentification of users. For example, 84% of users can be uniquely reidentified using only their preferences for eight movies that are not among the top 500 most popular movies. Due to the sparsity of user ratings, such reidentification can happen even with only a small amount of background knowledge about a target.

Considering that user ratings for recommendation are always sparse, some privacy protection is needed when user ratings are released. In this paper, we consider practical models for making recommendations based on anonymized ratings. We then introduce recommendation algorithms that fit these models and achieve a good balance between usefulness and preservation of privacy.

1.1 Related Works

Recently, there have been two lines of active research concerning the preservation of privacy in recommendation systems. The objective of one of these lines of study is to preserve the privacy of the user ratings from the entity who provides the recommendation. Canny has presented a privacy-preserving recommender system that uses cryptographic tools and assumes that users communicate with each other over a peer-to-peer network [2]. In this system, the ratings are homomorphically encrypted, and then the matrix is completed without decryption using singular value decomposition. Polat et al. presented a recommendation scheme in which user ratings are randomized by the users themselves, and the entity that collects the user ratings perform the recommendations based on the randomized user ratings [10]. Through experiments, they examined the trade-offs between privacy and prediction accuracy in their recommendation scheme. Parra-Arnau et al. introduced a recommendation system in which the users can suppress sensitive ratings [8]. They also discussed the optimal trade-off between the suppression of ratings and privacy, but they did not explicitly discuss the effect of the privacy protection mechanism on the prediction accuracy.

In the other main line of research, the objective is to preserve privacy when releasing user ratings to be used for recommendations. Parameswaran et al. presented a scheme that first obfuscates the user ratings and then distributes them. Their obfuscation scheme does not necessarily give a mathematical guarantee of anonymity, but their experimental results revealed that it does not degrade the prediction accuracy of the recommendations [7]. Chang et al. proposed an algorithm that k-anonymizes user ratings and then constructs a recommender system based on the anonymized ratings. Recognizing that the rating matrices are sparse, Chang et al.'s idea is to first make the rating matrix dense by complementing it using singular value decomposition, and then to perform clustering with the complemented user-item rating matrix so that each cluster contains at least k users [3]. Experimental results with Netflix data show that the loss of accuracy in the predictions is less with the predictive anonymization scheme than it is with the regular k-anonymization scheme.

1.2 Our Contribution

Assume a collector of user ratings distributes k-anonymized ratings to the public, such as considered in [7] and [3]. The objective of our study is to survey all recommender systems that use anonymized ratings and then to present item-based collaborative filtering recommendation algorithms for each of these models. In general, a system that provides recommendations needs to have some information associated with each of the users in order to personalize the recommendations. Such information is referred to as the *prediction input*. Existing privacy-preserving recommendation algorithms implicitly assume that the system can relate the user who obtains a recommendation to the user who provided the anonymized ratings. However, this is not always the case in practice, because there is no guarantee that the ratings of any individual user are contained in the anonymized rating matrix that the system possesses. Even if they are contained, the recommender cannot know this.

Considering the above, we will introduce models for providing recommendations that are based on anonymized rating matrices, and which are determined by several input parameters that relate the individual who provided the ratings with the one who obtains the recommendation. We then present a recommendation algorithm for each of these models. In some models, our experimental results show that the prediction accuracy decreases when there is a strong guarantee of anonymity. This indicates that there is a trade-off between anonymity and utility; this does not contradict the conclusions of [3]. On the other hand, for some of our models, we found through experiment that the prediction accuracy of the recommendations that were based on anonymized ratings was better than those based on non-anonymized ratings, provided that the non-anonymized ratings were used as the prediction inputs. This surprising result suggests that the guarantee of anonymity does not necessarily degrade the utility if the recommendation model is appropriately chosen.

The paper is organized as follows. Section 2 introduces notations utilized throughout the paper. Section 3 defines four available recommendation models based on anonymized ratings. Section 4 presents recommendation algorithms that use ratings anonymized by item-similarity based collaborative filtering using the models derived in Section 3. Section 5 reports the results of experiments carried out to evaluate the effects of anonymization on the prediction accuracy of recommendations. Section 6 concludes the paper.

2 Preliminaries

2.1 Recommendation System

Let $(r_{ui}) = \boldsymbol{R} \in \mathbb{R}^{n \times m}$ be a sparse rating matrix of n users and m items, where r_{ui} is the rating of user u for item i. We denote the rating matrix as $(\boldsymbol{r}_1\boldsymbol{r}_2 \dots \boldsymbol{r}_n)^T = \boldsymbol{R}$, where $\boldsymbol{r}_u \in \mathbb{R}^m$ contains the ratings of user u. $\boldsymbol{R}$ contains missing values. Ω denotes a set of user-item pairs (u, i) so that r_{ui} has a value in $\boldsymbol{R}$.

Let $f : U \times I \mapsto \boldsymbol{R}$ be a prediction function of ratings where $U = \{1, 2, \cdots, n\}$ and $I = \{1, 2, \cdots, m\}$ denote the sets of user and item identities, respectively. For $(u, i) \notin \Omega$, the prediction of the rating is given by $\hat{r}_{ui} = f(u, i)$. Similarity-based collaborative filtering [4, 13] and matrix completion via low-rank matrix approximation [9] are well known as methods for recommender systems.

To evaluate the prediction accuracy of the recommender system, we divided Ω into two disjoint sets: a training set Ω_{train} and a test set Ω_{test}. The set Ω_{train} is used to train the prediction function, and the set Ω_{test} is used to evaluate its accuracy at making predictions. In our experiments, the root-mean-square error (RMSE) evaluated with k-fold cross-validation was used to measure the prediction accuracy of f.

2.2 k-Anonymity and k-Anonymization of Rating Matrices

Let T be a database with n records. We assume each record contains information associated with a single individual and consists of the *identifier*, *quasi-identifiers* (QIs), and *sensitive attribute values*. The identifier is a unique key that identifies the individual. The QIs are user attributes, such as age or gender; the particular combination of QIs can identify a single individual with high probability. After the identifiers have been eliminated, T is said to be *k-anonymous* if, for any combination of the QIs that appear in T, there exist at least k individuals who have that particular combination of QIs [12, 14]. A set of records that share the same combination of QIs is referred to as an *equivalence class*. Thus, if the size of all the equivalence classes in T is greater than k, T is k-anonymous.

In a recommendation system, each record contains the rating values associated with a particular individual. The rating values are usually interpreted as sensitive attribute values because they are not user attributes. However, as already mentioned in the introductory section, in some cases, individuals can be uniquely identified by the ratings with high probability, particularly when T is sparse. Thus, we regard the rating values as QIs.

Let t_u be a record associated with user u. In our setting, a record represents a rating vector, $t_u = \boldsymbol{r}_u \in \mathbf{R}^m$, and the database table is given as $T = \{\boldsymbol{r}_u\}_{u=1}^{n}$. Following the definition of k-anonymity [12, 14], for any rating vector in T, if there exist at least $k-1$ other rating vectors having the same rating values, then the database is k-anonymous. Let $\widetilde{T} = \{(\widetilde{\boldsymbol{r}}_u, k_u)\}_{u=1}^{n'}$ represent the k-anonymized database of T where $\widetilde{\boldsymbol{r}}_u$ for $u = 1, \dots n'$ are distinct and $k_u = |\{\boldsymbol{r} \in \mathbf{R}^m | \boldsymbol{r} = \widetilde{\boldsymbol{r}}_u\}|$. Since $k_u \geq k$ holds for any $(\widetilde{\boldsymbol{r}}_u, k_u) \in \widetilde{T}$, $\widetilde{T}$ is k-anonymous.

Indices $1 \leq u \leq n$ of T and indices $1 \leq u \leq n'$ of $\widetilde{T}$ are referred to as the *user identity* and *anonymized identity*, respectively. The onto mapping from a user identity to an anonymized identity is defined by $\sigma : U \mapsto U'$, where U and U' are the domains of the user identities and the anonymized identities, respectively.

In recommendation systems, the number of items is large in general, which means that the rating vectors tend to have high dimensionality. Since the number of combinations of rating values exponentially increases with respect to the number of items, k-anonymization by means of generalization or suppression would seriously destroy the nature of the original rating values. For k-anonymization of high-dimensional numerical values, it is known that clustering or microaggregation preserves the utility of the original data. Here is an outline of k-anonymization by clustering: First, an algorithm clusters the vectors so that every cluster contains at least k vectors. Then, each vector is replaced by the prototype of the cluster to which that vector is assigned. Algorithms for k-anonymization by clustering include the one-pass K-means algorithm (OKA) [5] and r-gather clustering [1]. The recommendation models introduced in the following sections are not dependent on a specific anonymization method, but we used the OKA in our experiments. OKA is efficient in the sense that it completes after a one-pass scan of the entire data. Given a distance function that appropriately measures the difference between two data points, the algorithm first selects k datapoints uniformly at random, which are set as the cluster centers. The algorithm scans the remaining data points one by one; measures the distance between the data point and cluster centers, and assign the data point to the nearest cluster center.

3 Recommendation Models Based on Anonymized Ratings

In this section, we survey the available recommendation models that are based on anonymized ratings and then examine the risks of de-anonymization of the ratings.

3.1 Stakeholders

We first introduce the four stakeholders that occur in recommendations with anonymized ratings: the *rater*, the *rating collector*, the *recommender*, and the *user*.

The rater is the entity that gives the rating values to the rating collector. The rating collector (*collector* for short) is the entity that collects the rating values from the raters and constructs the sparse rating matrix $\mathbf{R}$. Then, if necessary, the rating collector anonymizes the rating matrix as $\widetilde{\mathbf{R}}$ and distributes it to the recommender. The recommender is the entity that obtains rating matrix $\mathbf{R}$ (or anonymized rating matrix $\widetilde{\mathbf{R}}$) from the collector and constructs the prediction function f. Then, upon request, the recommender provides recommendations to the users.

In regular recommendation systems, we expect that the recommender system can identify users as raters at prediction time if the raters contributed their ratings previously. However, in recommendations based on anonymized ratings, the recommender system cannot necessarily identify users as raters uniquely due to anonymization even if the users previously contributed their ratings. Identification of users at prediction time are discussed in Section 3.3 in detail.

3.2 Training Input

The rating matrix that the recommender obtains from the collector is called the *training input*. Two types of training inputs are considered in our models.

Case 1. Let U and U' be the set of raters and users, respectively. We assume $U' \subseteq U$ in regular recommendation systems. In Case 1, the prediction for any user can be personalized by using the ratings given by the user in the past. We represent the rating matrix in this case as $\boldsymbol{R}_+$ (Table 1, line 3). If the collector anonymizes the rating matrix before providing it to the recommender, we call this Case 1A. In Case 1A, the training input, i.e., the anonymized rating matrix, is represented as $\widetilde{\boldsymbol{R}}_+$ (Table 1, line 4).

Case 2. In Case 2, we assume $U \cap U' = \emptyset$. That is, none of the users who wish to obtain recommendations have previously given ratings to the collector. In this case, the predictions cannot be personalized based on past ratings. This situation is known as a *cold start*. The rating matrix, in this case, is represented by $\boldsymbol{R}_-$ (Table 1, line 5). If the collector provides the anonymized rating matrix $\widetilde{\boldsymbol{R}}_-$ of $\boldsymbol{R}_-$, the situation is referred to as Case 2A (Table 1, line 6).

If the collector is a company whose customer database is comprehensive (e.g., a railway or cell phone provider), and the recommender wishes to provide recommendation services based on anonymized ratings purchased from the collector, then the recommendation system can be modeled with Case 1A. However, if the ratings are collected from a limited segment of customers and are anonymized before distribution, then the recommendation system can be modeled with Case 2A; this is because the customer databases of the collector and the recommender are disjoint. Of course, there are intermediate situations between Case 1/1A and Case 2/2A that can be considered; however, we will not pursue these because they can be covered by extensions of the existing models.

3.3 Prediction Input and Models of Recommendation

In order to provide a prediction that is personalized for a particular user, the recommender needs to have information about the user. This information is referred to as *prediction input*. Let $\omega(u)$ be the prediction input for user u. Below, we list possible variations of the

Table 1: Recommendation models based on anonymized ratings

	training input	prediction input			
		user id u	user ratings $\boldsymbol{r}_u$	anonymized id $\sigma(u)$	$\emptyset$
Case 1	$\boldsymbol{R}_+$	Case1/REG	—	—	BASELINE
Case 1A	$\widetilde{\boldsymbol{R}}_+$	—	Case1A/UR	Case1A/AI	
Case 2	$\boldsymbol{R}_-$	—	Case2/UR	—	
Case 2A	$\widetilde{\boldsymbol{R}}_-$	—	Case2A/UR	—	

prediction inputs. Table 1 summarizes the relationships between the training input and prediction input for the different recommendation models.

1. User identity $\omega(u) = u$ as prediction input

   Regular recommendations based on non-anonymized ratings are modeled with this prediction input; users provide their identities (Case1/REG). In Case 2 and Case 2A, user identities are not contained in the training inputs. In Case 1A, the recommender cannot know the relationship between the users and the anonymous raters of the training inputs. Thus, in these cases, user identities cannot be used as prediction input.

2. Anonymous identity $\omega(u) = \sigma(u)$ as prediction input[1]

   Note that this prediction input can be used only in Case 1A because in Case 2A, we assumed the ratings of users are not contained in the training input. In Case 1A, if an anonymous identity is given to the recommender as a prediction input, the prediction is personalized not for the user, but for the anonymous identity that contains the user. Because the prediction is personalized for the k or more users having the anonymous identity, the effect of personalization can be weakened.

3. User ratings $\omega(u) = \boldsymbol{r}_u$ as prediction input

   In this case, users provide some of the ratings as prediction input. In Cases 1A, 2, and 2A, the recommender cannot connect the prediction inputs with the raters in the training inputs. However, if a user can independently provide rating values at prediction time, aside from the training inputs, the recommender can personalize the prediction for the user, based on the ratings provided (Case1A/UR, Case2/UR, and Case2A/UR). In Case1A/UR and Case2A/UR, the training inputs that the recommender obtains are anonymized while the users provide non-anonymized ratings at prediction time. If the users put more focus on the prediction accuracy of recommendation than on preserving anonymity, this decision is reasonable.

4. No prediction input: $\omega(u) = \emptyset$.

   Without prediction input associated with the users, the predictions cannot be personalized. In this case, the recommender can do nothing but to use the average ratings of the items (BASELINE).

[1] Recall that the onto mapping from a user identity to an anonymized identity is defined by $\sigma : U \mapsto U'$, where U and U' are the domains of the user identities and the anonymized identities, respectively.

3.4 Risk of Privacy Leakage

When making recommendations with anonymized ratings, two types of privacy risks should be considered. In the first case, when the anonymized ratings are used as the training input, the recommender may try to use the anonymized rating vectors to identify the raters. In the second case, when the recommender obtains the prediction input from the users, the recommender may try to use the vectors to identify the users.

In the first case, the privacy risk is dependent on the guarantee of anonymity of the training input. If it is k-anonymous, the probability with which the recommender can identify the rater is at most $1/k$.

Next, we consider the risks in the second case. For Case 2A, user ratings are not contained in the collection for any type of prediction input; there is thus no risk of re-identification. For Case 1A/AI, the recommender obtains $\widetilde{\boldsymbol{R}}_+$ as the training input and the anonymized identity $\omega(u) = \sigma(u)$ as the prediction input. However, what the recommender can infer from this is no more than what the recommender could estimate from $\widetilde{\boldsymbol{R}}_+$ alone. Thus, the probability with which the recommender can reidentify the rater is again $1/k$ at most.

On the other hand, for Case1A/UR, if the recommender obtains $\omega(u) = \boldsymbol{r}_u$ as the prediction input, the degree of anonymity can be decreased. For example, suppose the training input is a three-anonymous rating collection and Users 1, 2, and 3 all belong to the same anonymized identity. If the ratings of User 1, $\boldsymbol{r}_1$, are used as the prediction input, the recommender will be able to learn the anonymized identity to which User 1 belongs. The contribution of User 1 can then be removed from the rating vector associated with $\sigma(1)$, which is given as the cluster center of the three users. This means that the guarantee of k-anonymity has been degraded from k-anonymity to $(k-1)$-anonymity.

Note that the anonymity can be degraded even if no information is provided that is associated with User 2 or 3. For Case1A/UR, in order to guarantee k-anonymity after the prediction, the collector needs to anonymize the rating collection with some integer larger than k. Also, the number of user ratings that the recommender can obtain needs to be controlled so that the k-anonymity of the training input is not compromised.

4 Item-similarity Based Collaborative Filtering with Anonymized Ratings

In this section, we present recommendation algorithms that use ratings anonymized by item-similarity based collaborative filtering [13, 11], using the models derived in the previous section. These algorithms use training inputs to construct groups of similar items before generating personalized predictions. We first introduce item-similarity based collaborative filtering without anonymization and then present item-similarity based collaborative filtering with anonymization.

4.1 Case1/REG

The similarity of two items is defined to be the Pearson correlation coefficient of their ratings. Let U_i be the set of users who rate item i. Then, the average rating of item i and the correlation coefficient between item i and item j are given by

$$r_{*i} = \frac{1}{|U_i|}\sum_{u \in U_i} r_{ui}, \qquad s_{ij} = \frac{\sum_{\ell=1}^{m}(r_{\ell i} - r_{*i})(r_{\ell j} - r_{*j})}{\sqrt{\sum_{\ell=1}^{m}(r_{\ell i} - r_{*i})^2}\sqrt{\sum_{\ell=1}^{m}(r_{\ell j} - r_{*j})^2}}. \tag{1}$$

With these item similarities, the predicted rating of item i for user u is given by

$$f(u, i; \omega(u) = u) = r_{*i} + \frac{\sum_{\ell \in I_u} s_{i\ell}\,(r_{u\ell} - r_{*\ell})}{\sum_{\ell \in I_u} |s_{i\ell}|}, \tag{2}$$

where I_u is the set of items rated by user u.

4.2 Case2/UR

Let u' be a user who wishes to obtain a recommendation. In this case, the ratings of user u' are not contained in $\boldsymbol{R}_-$. Since $I_{u'} = \emptyset$ in this case, it cannot be predicted by using eq. 2. In order to personalize the predictions, users provide the $\boldsymbol{r}_{u'}$ of some of the items as the prediction input. Note that item similarities can be evaluated independently on users. Then, the prediction for u' can be given using $\boldsymbol{r}_{u'}$ as

$$f(u', i; \omega(u') = \mathbf{r}_{u'}) = r_{*i} + \frac{\sum_{\ell \in \bar{I}_{u'}} s_{i\ell}\,(r_{u'\ell} - r_{*\ell})}{\sum_{\ell \in \bar{I}_{u'}} |s_{i\ell}|}, \tag{3}$$

where $\bar{I}_{u'}$ is the set of items rated by user u' in $\boldsymbol{r}_{u'}$. Here, note that $\bar{I}_{u'}$ can be arbitrarily chosen by user u'.

4.3 Case1A/UR and Case2A/UR

We now describe the algorithm for Case1A/UR. The item similarities are evaluated from the anonymized training input $\widetilde{\boldsymbol{R}}_+ = (\tilde{r}_{ui})$ by

$$\tilde{s}_{ij} = \frac{\sum_{\ell=1}^{m}(\tilde{r}_{\ell i} - \tilde{r}_{*i})(\tilde{r}_{\ell j} - \tilde{r}_{*j})}{\sqrt{\sum_{\ell=1}^{m}(\tilde{r}_{\ell i} - \tilde{r}_{*i})^2}\sqrt{\sum_{\ell=1}^{m}(\tilde{r}_{\ell j} - \tilde{r}_{*j})^2}}, \tag{4}$$

where $\tilde{r}_{*i}$ is the average rating of item i evaluated with $\widetilde{\boldsymbol{R}}_+$.

Then, the prediction is made using the non-anonymized user ratings given as prediction inputs by

$$f(u, i; \omega(u) = \boldsymbol{r}_u) = \widetilde{r}_{*i} + \frac{\sum_{\ell \in \bar{I}_u} \widetilde{s}_{i\ell}\,(r_{u\ell} - \widetilde{r}_{*\ell})}{\sum_{\ell \in \bar{I}_u} |\widetilde{s}_{u\ell}|}. \tag{5}$$

Prediction by eq. 5 makes use of the average item ratings and the item similarities estimated from the anonymized rating matrix, whereas the prediction is made for a particular user. By giving user ratings as prediction inputs, the prediction can be personalized even when the rating matrix has been anonymized.

In the algorithm for Case2A/UR, the similarities can be obtained by using eq. 4; the predictions can be made with eq. 5 where $\bar{I}_u$ is replaced by $\bar{I}_{u'}$, as in eq. 3.

4.4 Case1A/AI

In this case, the similarities can be evaluated by eq. 5. The prediction of the rating of item i for user u becomes

$$f(u,i;\omega(u)=\sigma(u))=\widetilde{r}_{*i}+\frac{\sum_{\ell\in\widetilde{I}_{\sigma(u)}}\widetilde{s}_{i\ell}\left(\widetilde{r}_{\sigma(u)\ell}-\widetilde{r}_{*\ell}\right)}{\sum_{\ell\in\widetilde{I}_{\sigma(u)}}|\widetilde{s}_{i\ell}|}, \tag{6}$$

where $\widetilde{I}_{\sigma(u)}$ is the set of items rated by anonymous identity $\sigma(u)$ in the anonymized rating matrix $\widetilde{\boldsymbol{R}}_{+}$. Note that predictions made by this equation are personalized for an anonymous user identity $\sigma(u)$, not for a particular user u.

4.5 BASELINE

In the BASELINE model, the recommender does not have any information that can be used to personalize the prediction. Therefore, the prediction is the average of the anonymized ratings of the specified item: r_{*i} as evaluated by eq. 1.

In summary, Case1/REG is equivalent to [13]. Case2/UR is a well-known extension of Case1/REG for the cold-start setting. Case1A/AI is a model similar to [10] or [3]. Case1A/UR and Case2A/UR are models that, to the best of our knowledge, are newly introduced in this paper.

5 Experiments

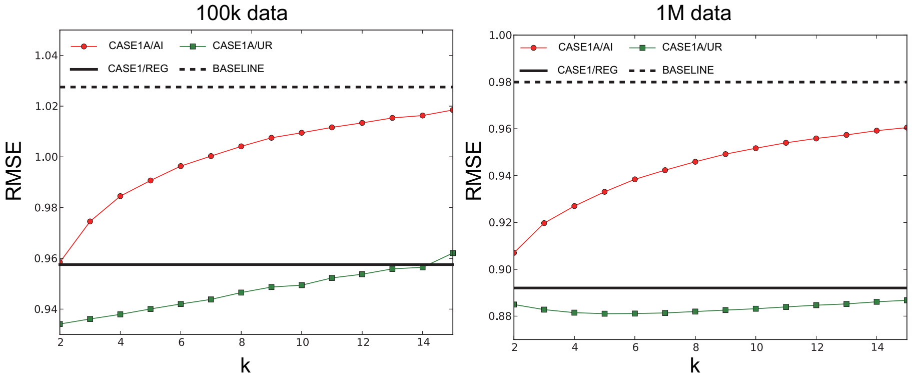
In this section, we report the results of experiments carried out in order to evaluate the effects of anonymization on the prediction accuracy of recommendations. For these experiments, we used the MovieLens datasets [11]. The 100k dataset contains 100,000 ratings of 1682 items given by 943 users. The 1M dataset contains about a million ratings of 3883 items given by 6040 users.

5.1 Experiments for Case 1 and Case1A

5.2 Settings

For Case 1 and Case 1A, we assumed that all the users who received personalized predictions were contained in the set of raters, i.e., $U' \subseteq U$. For Case1/REG, we uniformly chose 80% of the ratings from $\boldsymbol{R}$ at random and used them as the training input. The remaining ratings were used for the test data to evaluate the RMSE.

For Case1A/AI and Case1A/UR, 80% of the ratings from $\boldsymbol{R}_{+}$ were chosen uniformly at random, and then these ratings were k-anonymized using the OKA method [5]; these were used as the training input. The remaining 20% of the ratings were used as the test data. For Case1A/UR, 20% of the non-anonymized ratings $\boldsymbol{r}_u$, chosen uniformly at random from the training data, were used for the prediction. In the experiments, the RMSE of the recommendations with these models were measured while varying the anonymity parameter as $k=2,3,\ldots,15$.



(a) Change of RMSE w.r.t. k, 100k data

(b) Change of RMSE w.r.t. k, 1M data

Figure 1: Change of RMSE w.r.t. k for Case1/REG, Case1/UR, Case1/AI, and the BASELINE model. Left: 100k data, Right: 1M data.

5.3 Results

The change in the RMSE w.r.t. the anonymity parameter k for Case1/REG, Case1A/UR, Case1A/AI, and the BASELINE are shown in Fig. 1 . The RMSEs of Case1/REG are less than those of Case1/AI for all evaluated values of k. For Case1/AI with a larger k, the anonymous identities contain a larger number of individuals; this makes the target of the personalization vague, and the prediction accuracy deteriorates. This indicates that there is a trade-off between utility and privacy for Case1/AI.

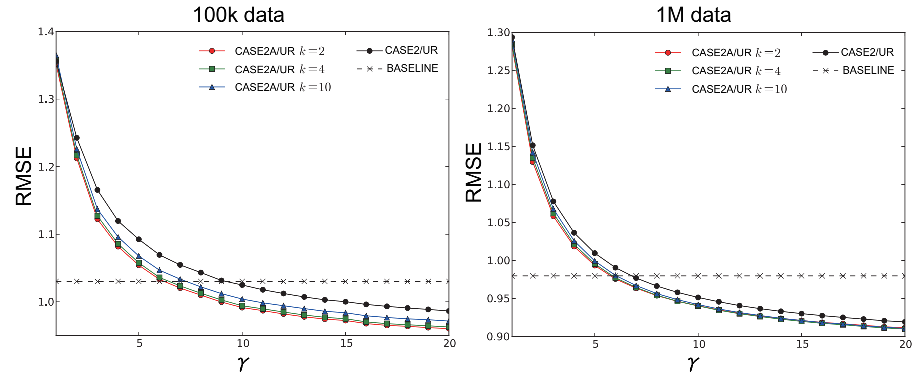
In Case1/UR, the RMSE deteriorates as k increases. However, surprisingly, Case1A/UR achieves a prediction accuracy that is better than that of Case1/REG for all evaluated values of k. This behavior is further discussed in Section 5.7. We must bear in mind that the anonymity in Case1A/UR can be weakened if the user ratings are given to the recommender as prediction input, as discussed in Section 3.4. Because the levels of anonymity achieved for Case1A/AI and Case1A/UR are not equivalent, the RMSE of Case1A/AI and Case1A/UR are not directly comparable.

For Case 1A/UR, the results are significantly different between the 1M dataset and the 100k dataset. With the 1M data, the RMSE slightly improves (decreases) as k increases in the range of $2 \leq k \leq 5$, then the RMSE increases as k increases in the range of $5 < k \leq 15$. This behavior is further discussed (and further experimental results are presented) in Section 5.7.

5.4 Experiments for Case 2 and Case 2A

5.5 Settings

For Case 2 and Case 2A, we assumed a cold start; none of the users who received personalized predictions were contained in the set of raters, i.e., $U' \cap U = \emptyset$. In Case2/UR, we uniformly chose 80% of the users in U at random, and their ratings were used as the training inputs. The ratings of the rest of the users were used for the test data to evaluate the RMSE. For the remaining 20% of the users, γ ratings were used as their prediction inputs and the rest of ratings were used as the test data to evaluate the RMSE. To evaluate



(a) Change of RMSE w.r.t. γ, 100k data

(b) Change of RMSE w.r.t. γ, 1M data

Figure 2: Change of RMSE w.r.t. γ, the number of user ratings used as the prediction input, for Case 2 and Case 2A. Case2/REG and Case2A/UR for $k = 2, 4$, and 10, and the BASELINE are compared. Left: 100k data, Right: 1M data.

the RMSE, 20 γ ratings were chosen uniformly at random to be used as prediction inputs; the RMSE averaged over these 20 trials was used as the RMSE. In these experiments, the number of ratings used as the prediction inputs, γ, was varied from 1 to 20; the anonymity parameter was set to $k = 2, 4, 10$. With these settings, we compared the changes in the RMSE for the BASELINE and the Case2/UR and Case2A/UR models w.r.t. γ.

5.6 Results

Figure 2 shows the changes in the RMSE of the BASELINE, Case2/REG, and Case2A/UR w.r.t. γ. Case2A/UR had a better the prediction accuracy than did Case2/UR, for all values of k and γ. From these results, we can see that item-similarity matrices evaluated with anonymized ratings gave better predictions for both Case 1A/UR and Case 2A/UR. Furthermore, we can see that the RMSEs are not greatly affected by the degree of anonymity k for Case2A/UR. Thus, if $k \leq 10$, anonymization of the ratings does not damage the recommendations.

Another important observation is that the RMSE improves as γ, the number of ratings used in the prediction input, increases. This is because the prediction can be better personalized when there are more user ratings used as prediction inputs. The RMSE of Case2A/UR is better than that of the BASELINE when $\gamma \geq 6$ in the 1M dataset and $\gamma \geq 8$ in the 100k dataset. Thus, a practical solution for giving recommendations with anonymized ratings is for the users to rate only a very small number out of thousands of items. Note that the ratings used for the prediction inputs can be arbitrarily chosen by the users. If the users provide ratings of popular items that are rated by a large number of users, privacy leakage caused by giving prediction inputs to the recommender might be ignored.

5.7 Analysis of Similarity Matrix Evaluated with Anonymized Ratings

From the experimental results in the previous subsections, we observe that the prediction accuracy of recommendations based on anonymized ratings can be better than those based on non-anonymized ratings in some settings; thus contradicts our intuition. In this subsection, we discuss the reasons for these useful but counterintuitive results.

The rating matrices used for the recommendations are usually quite sparse. In the Movielens dataset, the sparsity is 6.3% in the 100k dataset and 4.2% in the 1M dataset. In our experiments, clustering (OKA) was used for anonymization, and the anonymized user ratings were set to the average of the user ratings that belonged to the identical anonymous identities. If some but not all users who belonged to an anonymous identity rated an item, when the ratings were anonymized, the ratings of all members of the identity were set to the average rating of those users who did rate the item. Because of this manipulation, as k increases, a larger number of unrated elements were complemented with the average ratings. Thus, the anonymized rating matrix became denser. On the one hand, if the rating matrix became denser because k was larger, the item similarities can be estimated with a larger number of ratings, and this might estimate the similarities more precise. On the other hand, when k is larger, the number of users contained in a single anonymous identity also becomes larger. Thus, from the perspective of the users, the ratings of the anonymized identity do not accurately reflect their individual ratings, and this might cause the prediction accuracy to deteriorate. Thus, there exists a dilemma in setting the value of k.

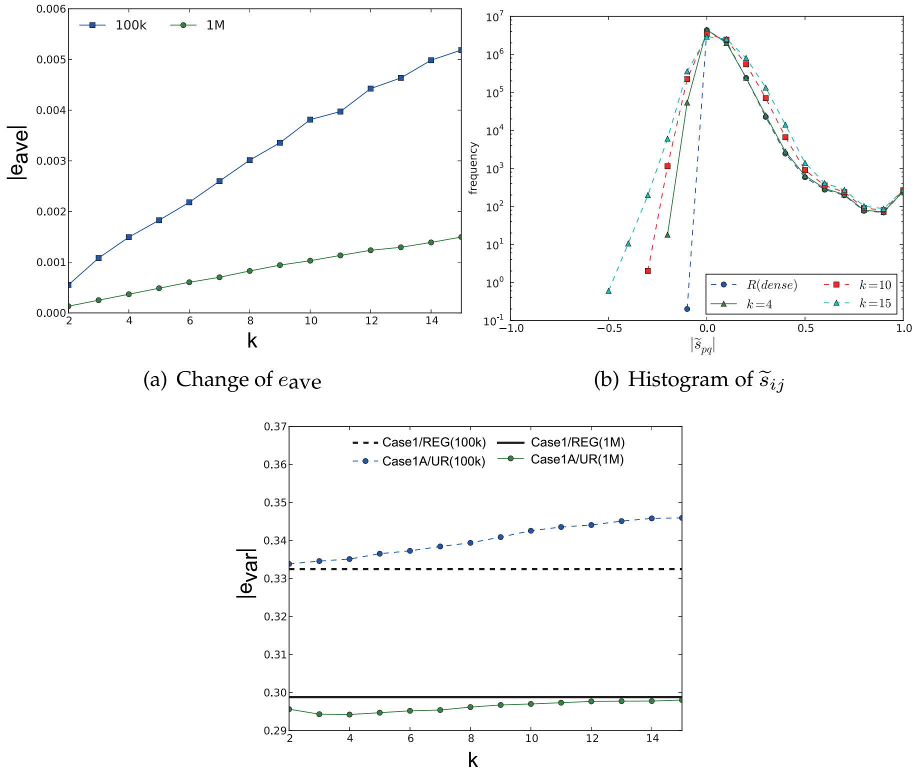
When recommendations are based on item similarities, predictions are made based on the average ratings of the items and the average of the user ratings weighted by the item similarities. In order to see how k affects the prediction accuracy, we examined the behaviors of the average ratings of the items and the item similarities w.r.t. changes in k.

Figure 3(a) shows the changes in the item ratings $e_{\text{avg}} = \frac{1}{|M|}\sum_{i=1}^{M} |r_{*i} - \tilde{r}_{*i}|$. In both the 100k dataset and the 1M dataset, e_{avg} becomes larger as k becomes larger; however, the absolute value of the error is kept within 0.01: the anonymity degree k does not greatly affect the average ratings of the items.

Figure 3(b) shows the histogram of the item similarities $\widetilde{s}_{ij}$ with different k values. From the figure, we can see that the positive similarities frequently appear in the non-anonymized item-similarity matrix, while the negative similarities do not. This tendency changes with larger k. More precisely, by complementing the sparse rating matrix by anonymization with larger k, the frequency of negative similarities increases, whereas that of the positive similarities does not change. The estimation of item similarities between two items by using eq. 1 becomes more accurate as the number of users who rate both of the items increases. Thus, the results shown in Fig. 3(b) indicate that anonymization enhances the evaluation of the item similarities between items that are not similar to each other. We should bear in mind that the similarity estimation with anonymized rating matrices does not necessarily provide "better" similarities compared to those estimated with non-anonymized rating matrices.

Finally, we consider the variance e_{var} of the prediction error $|r_{ui} - \hat{r}_{ui}|$, where $\hat{r}_{ui}$ is the prediction of the rating of item i for user u. Lower variance means better accuracy in general. Also, it represents the robustness of the prediction; a recommender with a lower variance of the prediction error gives more robust predictions. In Fig. 3(c), the changes of variance e_{var} w.r.t. k are shown for Case1/REG and Case1A/UR for 100k data and 1M data.

First, the variance in Case1A/UR is higher than that in Case1/REG for 100k data while we have the opposite result in 1M data. This indicates that recommendation with anonymized ratings can be more robust if we have a large-scale training input. The results show that anonymization with larger k causes a greater variance of the prediction error. However, for Case1A/UR, the variance decreases from $k = 2$ to 4, then it increases. This result



(a) Change of e_{ave}

(b) Histogram of $\widetilde{s}_{ij}$

(c) Change of e_{var}

Figure 3: (a) Change of e_{ave}, the difference between the average ratings of items in $\boldsymbol{R}$ and $\widetilde{\boldsymbol{R}}$, w.r.t. k; (b) Histogram of $\widetilde{s}_{ij}$; (c) Change of e_{var}, the variance of the difference between rating r_{ij} and the predicted rating $\hat{r}_{ij}$, w.r.t. k.

indicates that not only the prediction error but also the variance of the prediction error can be reduced by carefully tuning k in Case1A/UR.

6 Conclusion

In this paper, we considered recommender systems that take anonymized ratings as training inputs and give recommendations in various settings at prediction time. Our experimental results show that item-based collaborative filtering trained with anonymized ratings can give better prediction accuracy than that trained with non-anonymized ratings, when the users show 5–10 (non-anonymized) ratings to the recommender. We also observed that, by reducing the sparsity of the rating matrix by anonymization, the variance of the prediction error could be reduced if k, the anonymization parameter, is appropriately tuned. These surprising results indicate that privacy protection does not necessarily degrade the usefulness of recommendations. Our future work is to expand our models to other recommendation algorithms, including matrix factorization.

Acknowledgements

We thank anonymous reviewers for insightful reviews and discussions. We thank Prof. Vicenc Torra for important comments that greatly improved the manuscript. The research is partly supported by JST CREST Grant Number JPMJCR1302 and JSPS Grand-in-Aid for Scientific Research No. 16H02864.

References

[1] Gagan Aggarwal, Tomás Feder, Krishnaram Kenthapadi, Samir Khuller, Rina Panigrahy, Dilys Thomas, and An Zhu. Achieving Anonymity via Clustering. In *Proc. of the 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, page 153, New York, NY, USA, June 2006. ACM Press.

[2] John Canny. Collaborative filtering with privacy. In *Security and Privacy, 2002. Proceedings. 2002 IEEE Symposium on*, pages 45–57. IEEE, 2002.

[3] Chih-Cheng Chang, Brian Thompson, Hui Wendy Wang, and Danfeng Yao. Towards publishing recommendation data with predictive anonymization. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, pages 24–35. ACM, 2010.

[4] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237. ACM, 1999.

[5] Jun-Lin Lin and Meng-Cheng Wei. An Efficient Clustering Method for k-Anonymization. In *Proc. of the 2008 International Workshop on Privacy and Anonymity in Information Society - PAIS '08*, page 46, New York, New York, USA, March 2008. ACM Press.

[6] Arvind Narayanan and Vitaly Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy*, pages 111–125. IEEE, May 2008.

[7] Rupa Parameswaran and Douglas M Blough. Privacy preserving collaborative filtering using data obfuscation. In *Granular Computing, 2007. GRC 2007. IEEE International Conference on*, pages 380–380. IEEE, 2007.

[8] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. A privacy-protecting architecture for recommendation systems via the suppression of ratings. *Int. J. Secur., Appl., Sci., Eng. Res. Supp. Soc.(IJSIA)*, 6(2):61–80, 2012.

[9] Arkadiusz Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of KDD cup and workshop*, volume 2007, pages 5–8, 2007.

[10] Huseyin Polat and Wenliang Du. Privacy-preserving collaborative filtering using randomized perturbation techniques. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 625–628. IEEE, 2003.

[11] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proc. of the 1994 ACM Conference on Computer Supported Cooperative Work*, 1994.

[12] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, Technical report, SRI International, 1998.

[13] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. Item-based Collaborative Filtering Recommendation Algorithms. In *Proc. of the 10th International Conference on World Wide Web*, pages 285–295, New York, NY, USA, April 2001. ACM Press.

[14] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.