Expert Systems with Applications 39 (2012) 5033–5042

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# The impact of data obfuscation on the accuracy of collaborative filtering

Shlomo Berkovsky [a,*], Tsvi Kuflik [b], Francesco Ricci [c]

[a] Information and Communication Technologies Centre, CSIRO, Marsfield, Australia
[b] Department of Information Systems, The University of Haifa, Haifa, Israel
[c] Faculty of Computer Science, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy

ARTICLE INFO

*Keywords:*
Collaborative filtering
Recommender systems
Accuracy
Data obfuscation

ABSTRACT

Collaborative filtering (CF) is a widely-used technique for generating personalized recommendations. CF systems are typically based on a central storage of user profiles, i.e., the ratings given by users to items. Such centralized storage introduces potential privacy breach, since all the user profiles may be accessible by untrusted parties when breaking the access control of the centralized system. Hence, recent studies have focused on enhancing the privacy of CF users by distributing their user profiles across multiple repositories and obfuscating the user profiles to partially hide the actual user ratings. This work combines these two techniques and investigates the unavoidable side effect of data obfuscation: the reduction of the accuracy of the generated CF predictions. The evaluation, which was conducted using three different datasets, shows that considerable parts of the user profiles can be modified without observing a substantial decrease of the CF prediction accuracy. The evaluation also indicates what parts of the user profiles are required for generating accurate CF predictions. In addition, we conducted an exploratory user study that reveals positive attitude of users towards the data obfuscation.

Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

Collaborative filtering (CF) (Herlocker, Konstan, Borchers, & Riedl, 1999) is one of the most widely-used recommendation techniques. CF generates personalized recommendations based on the prediction of how much a user may like not yet examined items. The recommendation generation process is underpinned by the assumption that users who agreed in the past, i.e., whose opinions correlated in the past, will also agree in the future (Shardanand & Maes, 1995). The input for CF is a matrix that contains the user profiles represented by lists of user ratings for a set of items. To generate a prediction for an item, CF first computes the degrees of similarity between the *active user*, whose preferences are being predicted, and each one of the other users. Then, it selects a *neighborhood* of users having the highest degree of similarity with the active user. Finally, the prediction is generated by computing the weighted average of the neighbors' ratings for the item. Normally, the recommended items for the active user are those with highest predicted ratings.

However, personalization inherently brings to the fore the privacy challenge (Brier, 1997). Dealing with the user profiles implies that personal and potentially sensitive user data are collected and stored by CF recommender systems. Certain systems may violate the users' privacy by misusing data for their own benefits. As a result, users concerned about this misuse may refrain from using a system in order to minimize potential exposure of their data (Cranor, Reagle, & Ackerman, 1999). Privacy hazards for recommender systems are aggravated by the fact that the generation of recommendations requires large amounts of personal data, as the accuracy of CF predictions is correlated with the number of similar users and the number of ratings in their profiles (Sarwar, Karypis, Konstan, & Riedl, 2000). Hence, there is a trade-off between the accuracy of recommendations provided to users and the degree of their privacy.

This has triggered growing research in recommender systems. Canny proposed to enhance user privacy by a decentralized storage of user profiles (Canny, 2002). The users were grouped into communities, which represented the preferences of the underlying users as a whole and did not expose preferences of individual users. Alternatively, Polat and Du enhanced user privacy by adding uncertainty to the data stored in the user profiles (Polat & Du, 2005). This was accomplished by using randomized data obfuscation techniques, which modified the data in the user profiles. Hence, if exposed, the data disclosed the obfuscated rather than the real user preferences.

This work elaborates on the idea of combining the above two techniques, as initially discussed in Berkovsky, Eytani, Kuflik, and Ricci (2007). In a nutshell, we propose to (1) substitute the commonly used centralized CF systems with decentralized ones, and (2) add a degree of uncertainty by obfuscating the ratings in the

* Corresponding author.
*E-mail addresses:* shlomo.berkovsky@csiro.au (S. Berkovsky), tsvikak@is.haifa.ac.il (T. Kuflik), fricci@unibz.it (F. Ricci).

0957-4174/$ - see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.11.037

user profiles. In this setting, users participate in a decentralized CF system in the following way. The users maintain their profiles and independently obfuscate them. Predictions are requested by an active user through exposing their profile and sending prediction request to other users. Those users, who decide to respond to the request, also expose parts of their profile (rating for the requested item) and send it to the active user together with the computed degree of user-to-user similarity. Then, the active user collects the received responses, selects the neighborhood of the most similar users, and aggregates their ratings for the prediction generation.

In this setting, users possess a better control over their data, since they can decide when and how much data to expose. Moreover, they can select which part of their profile should be obfuscated and control the obfuscation parameters. However, this approach may have negative impact on the accuracy of the generated CF predictions, as the obfuscated ratings may differ from the real ones, which may lead to inaccurate similarity computation, neighborhood selection, and, as a result, to inaccurate predictions and recommendations. Hence, the impact of data obfuscation on the accuracy of CF recommendations is the primary focus of this work.

In the experimental part of this work, we evaluated the accuracy of the distributed CF with data obfuscation using three publicly available datasets: Jester (Goldberg, Roeder, Gupta, & Perkins, 2001), MovieLens (Herlocker et al., 1999), and EachMovie (McJones, 1997). The evaluation demonstrated that users can obfuscate large parts of their profiles without significantly decreasing the accuracy of the generated predictions. This raised a question regarding the contribution of various ratings in the user profiles to the accuracy of CF recommendations. Hence, additional experiments, aimed at analyzing the impact of data obfuscation applied to different part of user profiles, were conducted. The results obtained for all three datasets indicate that the accuracy of the recommendations is affected by *extreme ratings* having very positive or very negative values stronger than by moderate ratings. In fact, these ratings are also the most sensitive user data, as was ascertained by a study of 117 participants that showed that users perceive their extreme ratings to be more sensitive than moderate ratings.

Hence, the contribution of this work is threefold. Firstly, we propose and evaluate several data obfuscation policies and their impact on the accuracy of the generated CF predictions. Secondly, we compare the impact of data obfuscation applied to extreme and moderate ratings in the user profiles on the accuracy of the generated predictions. Thirdly, we assess the attitude of users towards the data obfuscation and their perception of the sensitivity of ratings.

The rest of the paper is organized as follows. Section 2 discusses related work on distributed CF and user profile obfuscation. Section 3 presents experimental results evaluating the proposed data obfuscation techniques. Section 4 presents the user study. Finally, Section 5 concludes the paper and outlines our future research directions.

## 2. Distributed collaborative filtering with data obfuscation

### 2.1. Related work

Centralized CF introduces a severe privacy threat, as personal information collected by service providers can potentially be acquired by untrusted parties. Thus, users may refrain from rating items and revealing their preferences due to privacy concerns (Cranor et al., 1999). For example, a survey showed that 83% of users are more than marginally concerned about protecting their privacy and 90% of users are concerned about misuse of their information (Ackerman, Cranor, & Reagle, 1999). Hence, generating accurate CF recommendations without compromising user privacy is an important challenge.

Three main threats that may hamper proper functioning of a CF recommender system are discussed in Lam, Frankowski, and Riedl (2006). *Exposure* refers to undesired access to a user's personal information by untrusted parties, who are not supposed to have this access. *Bias* refers to manipulation of user profiles in order to generate inappropriate recommendations, i.e., to increase or decrease visibility of certain items. *Sabotage* refers to intentional reduction in the functionality of a system, such as service denial or malfunctioning. While the latter two can be considered as functional threats that hamper the service of the system, the first threat is a clear privacy breach leading to unwanted disclosure of personal data. In this work, we focus on the privacy breach only and evaluate the impact of data obfuscation (as a means to address this breach) on the accuracy of the generated recommendations.

Distributed storage of user profiles partially mitigates the privacy breach, as the attacker needs to violate multiple repositories when attacking a decentralized system. The way of generating CF predictions in a distributed setting was outlined in Tveit (2001). There, the communication between a group of mobile users exploited a flooding-based routing mechanism, which increased the communication overheads. The PocketLens project (Miller, Konstan, & Riedl, 2004) compared five centralized and decentralized CF architectures and showed that the performance of a decentralized CF is close to that of a centralized one. Another approach to distributed CF eliminating the use of central servers was presented in Olsson (1998). There, active users initiated recommendation queries by sending parts of their profiles. Other users responded to the query and sent their information to the active user. However, this required transferring the user profiles and potentially creating more privacy breaches.

Another approach for a distributed CF was proposed in Canny (2002). Individual users controlled their private data and were grouped into communities, which represented public aggregation of their ratings. This allowed personalized CF predictions to be generated by exposing the aggregated community data without exposing the data of individual users. However, this approach required an a priori formation of communities, which may be a limitation in highly dynamic environments.

An alternative research direction focused on enhancing privacy of CF through data modification. Several data modification methods were discussed in Ioannidis, Grama, and Atallah (2002). These include encryption (Agrawal, Kiernan, Srikant, & Xu, 2004), access-control policies (Sandhu, Coyne, Feinstein, & Youman, 1996), data randomization (Agrawal et al., 2004), anonymization (Klösgen, 1995), and *k*-anonymization (Sweeney, 2002). These methods differ in the degree of privacy they provide and the extent to which they preserve the utility of the data, i.e., the accuracy of CF predictions. For instance, although encryption guarantees that nothing can be deduced from the modified ratings, it makes them practically useless for the active user. Generalization techniques, such as *k*-anonymization, preserve some resemblance between the original and the modified ratings, but allow an attacker to learn about the respondent through multiple attacks. In this work, we consider methods in which the modified ratings are independent of the original ones, i.e., every single modified rating discloses no information about the original rating.

Focusing on CF research, (Polat & Du, 2005) examined a centralized CF in which, before disclosing their profiles and sending them to the central server, users obfuscate their profiles. Hence, if the server was attacked and data exposed, an attacker would obtain only the obfuscated data. However, predefined obfuscation methods still allowed the attacker to recover the original ratings. In order to overcome the observed limitations, a new two-way

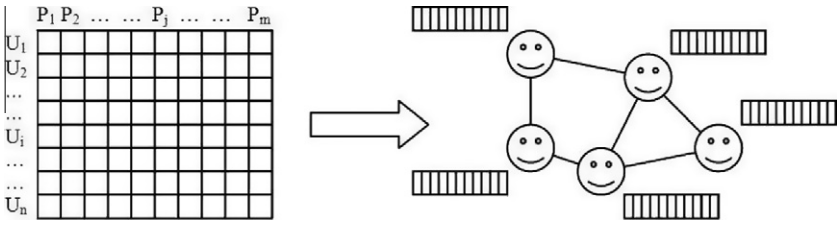


**Fig. 1.** Centralized vs. decentralized storage of the user profiles.

technique was suggested (Zhang, Ford, & Makedon, 2006). The users and the server agreed on a disclosure measure, and then the server sent to the user guidelines for modifying the ratings before sending them to the server.

Another framework for privacy protection through obfuscating sensitive information was proposed in Parameswaran and Blough (2007). This framework intended to overcome the cold start problem of CF, while preserving the users' privacy and allowing information sharing between service providers. Service providers obfuscated the data and sent them to a central server, which integrated the partial data, generated predictions, and sent them to the service providers. Alternatively, data modification can be conducted by the users. For example (Shokri, Pedarsani, Theodorakopoulos, & Hubaux, 2009) proposed to augment the user profile with profiles of similar users before sending the data to the centralized server. Thus, users keep their offline profiles and occasionally send it to the server, where the profiles are stored.

All these works demonstrate the interest in data obfuscation as a simple and intuitive means for privacy preservation. While it was shown that obfuscation may enhance user privacy, it is also important to investigate the impact of obfuscation on the accuracy of the generated predictions, which is the focus of this work.

### 2.2. Privacy-enhanced collaborative filtering with data obfuscation

This section elaborates on the prediction generation using a distributed set of users with possibly obfuscated profiles. In this work, we consider a decentralized organization of users, in which the users store and maintain their profiles. In particular, they can decide about the data obfuscation parameters: how many ratings to obfuscate, which ratings to obfuscate, how to obfuscate them, and which recommendation requests to respond to. Hence, traditional centralized CF matrix of user ratings for items is substituted by a virtual decentralized matrix. The rows of the matrix, i.e., the user profiles, are stored by the users. Fig. 1 illustrates the virtual distribution (right) of the centralized ratings matrix (left). The users are connected using a decentralized communication platform with no single point of management or failure.[1]

CF prediction generation process consists of the following stages:

- The active user sends a request for recommendation for the target item. For this, the active user sends their own profile and requests other users to participate in the distributed prediction generation.
- Every user receiving the request decides whether to respond. If decided to respond, they apply the agreed upon user-to-user similarity metric, compute the degree of similarity between the profile of the active user and their own profile, and send the degree of similarity degree and their rating for the target item to the active user.
- The active user selects the set of nearest neighbors. The neighbors can either be $K$ users having the highest degree of similarity with the active user, or all users whose similarity is above a threshold $\beta$.
- Finally, the active user computes the predicted ratings for the target item by aggregating the ratings of the neighbors on the target item in a weighted manner according to the users' similarity degree.

Note that the obfuscation is applied only by the responding users, i.e., the profile of the active user remains unchanged in order to increase the chances of getting accurate recommendations. Due to the same reason, the rating of the respondent for the target item is not obfuscated either.

In this distributed CF process, user information may be exposed in three cases. Firstly, when the profile of the active user is sent to other users as part of the recommendation request. Secondly, when the respondents send the degree of their similarity to the active user. Thirdly, when the respondents send their rating for the target item. It is worth to note that privacy concerns of users depend on their role in the CF process: the active user or the respondent. Unlike the active user, the respondent does not benefit from a response. Thus, the privacy concerns of respondents are the ones that threaten the livelihood of a CF system, especially if they reveal parts of their profile. Hence, this work focuses on the respondent's privacy breach and data obfuscation as a means to mitigate it.

We investigate three data obfuscation strategies:

- *Default obfuscation (x)* – substitute real ratings in the user profile with a predefined value $x$.
- *Uniform random obfuscation* – substitute real ratings in the user profile with randomly chosen values within the range of ratings in the dataset.
- *Distribution-based obfuscation* – substitute real ratings in the user profile with values reflecting the distribution of ratings in the dataset.

In the experimental section, we evaluate these strategies by obfuscating a portion of ratings in the user profiles and observing the impact of this obfuscation on the accuracy of the generated CF predictions. In practice, we first obfuscated the data and then let the users participate in the CF recommendation process. Thus, every user computed user-to-user similarity and responded to all incoming requests using the same obfuscated data.

### 2.3. Extreme ratings

Prior research suggested that not all the ratings are equally important in CF. For example, (Shardanand & Maes, 1995) argues that the accuracy of CF is crucial when predicting extreme, i.e., very high or very low ratings. That is, achieving accurate predictions for the best (or the worst) items is most important, since these

[1] Technical details of the platform are omitted. Readers are referred to surveys on decentralized technologies (Androutsellis-Theotokis & Spinellis, 2004; Milojicic et al., 2002).

**Table 1**
Rating obfuscation and prediction experiments.

| | Predicting extreme | Predicting all |
|---|---|---|
| Obfuscating extreme | Section 3.4 | Section 3.5 |
| Obfuscating all | Section 3.3 | Section 3.2 |

items should necessarily (or never) be recommended. Similarly, (Pennock, Horvitz, Lawrence, & Giles, 2000) focuses on the evaluation of the extreme rating predictions, motivating this by a similar assumption that users are mostly interested in recommendations for items they might love and warnings to avoid items they might hate, but not in recommendations for items they may moderately like.

Similarly, in this work we evaluate separately the importance of extreme ratings for data obfuscation. Hence, the obfuscation policies are applied to two groups of ratings: (1) *overall ratings* – all the available ratings in the dataset, and (2) *extreme ratings* – extremely positive or extremely negative ratings only (will be defined later). Similarly, we measure the impact of data obfuscation on the accuracy of CF predictions for two groups of ratings: (1) *overall predictions* – predictions of all the available ratings, and (2) *extreme predictions* – predictions of extremely positive or extremely negative ratings only (will also be defined later).

## 3. Experimental evaluation

This section presents the evaluation of the impact of data obfuscation on the accuracy of the generated CF predictions. The experiments reflect two groups of ratings and predictions: overall and extreme. The flow of the evaluation is summarized in Table 1: rows represent the groups of ratings that are obfuscated and columns represent the groups of ratings for which the predictions are generated and the accuracy is evaluated. The content of the table shows the sub-sections in which the corresponding evaluation is presented.

### 3.1. Experimental setting

For the evaluation, a decentralized setting of users was simulated by a multi-threaded implementation, such that each user was represented by a thread. The predictions were computed in a way presented in Section 2.2. The similarity of users was computed using the Cosine Similarity metric and $K = 10$ users having the highest similarity degree were included in the neighborhood.

To provide a solid evidence, the evaluation included three CF datasets: Jester (Goldberg et al., 2001), MovieLens (Herlocker et al., 1999), and EachMovie (McJones, 1997). Table 2 summarizes statistical properties of the datasets: number of users and items, range of ratings, number of ratings, average number of items rated by each user, density of the dataset, average and variance of ratings, and MAE of non-personalized predictions (average difference between the observed and average item rating).

To examine the impact of obfuscating and predicting extreme ratings, smaller *extreme datasets* containing a higher portion of extreme ratings were generated. For this, we extracted the ratings of extreme users, i.e., users having more than 33% of extreme ratings in their profiles. For user $u$ whose average rating is $av(u)$ and variance is $var(u)$, ratings above $av(u) + 1.5var(u)$ or below $av(u) - 1.5var(u)$ are considered as extreme.[2] For example, if $av(u) = 0.6$ and $var(u) = 0.2$, then ratings lower than 0.3 and higher than 0.9 are regarded as extreme. If $u$ rated 120 items and more than 40 ratings were extreme, then the data of this user are extracted into the extreme dataset. Table 3 summarizes statistical properties of the extreme datasets (columns are similar to Table 2).

Note that not all the ratings in the extreme datasets are necessarily extreme. More precisely, the extreme datasets contain users, who provided a high number of extreme ratings rather than extreme ratings only. The latter dataset would have required us to discard moderate ratings and create an unrealistic dataset. However, a comparison of the variance of ratings in the original and extreme datasets shows that the number of extreme ratings in the extreme datasets is considerably higher than in the original datasets. Fig. 2 shows the distribution of ratings in the datasets. The horizontal axis stands for the value (or range) of ratings and the vertical for the relative number of such ratings. Two distributions are shown: the light bars represent the distribution in the original dataset and the dark – in the extreme dataset. As can be seen, the distribution of ratings in the original datasets is bell-curve-like, whereas in the extreme datasets the bell-curve is inverted.

Three obfuscation strategies presented in Section 2.2 were instantiated by five specific policies:

- *Positive* – substitute a rating with the maximal rating in the dataset: 10 for Jester and 5 for MovieLens and EachMovie.
- *Negative* – substitute a rating with the minimal rating in the dataset: −10 for Jester, 1 for MovieLens, and 0 for EachMovie.
- *Neutral* – substitute a rating with the average of the maximal and minimal ratings in the dataset: 0 for Jester, 3 for MovieLens, and 0.5 for EachMovie.
- *Random* – substitute a rating with a random value in the range of ratings in the dataset: [−10, 10] for Jester, [1,5] for MovieLens, and [0,5] for EachMovie.
- *Distribution* – substitute a rating with a random value, such that the distribution of the substituted ratings reflects the distribution of original ratings in the dataset.

The *positive*, *negative*, and *neutral* policies are instances of the *default* strategy. The *random* policy is an instance of the *uniform* strategy, as the substituted ratings are chosen randomly in the range of ratings in the dataset. The *distribution* policy is the *distribution-based* strategy, as the substituted ratings reflect the distribution of ratings in the dataset. The relative number of obfuscated ratings in the user profile is referred to in this work as the *obfuscation rate*.

The accuracy of the generated predictions was measured using the normalized Mean Absolute Error (MAE) (Herlocker, Konstan, Terveen, & Riedl, 2004), a widely-used CF predictive accuracy measure. MAE was computed by:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - r_i|}{NR} \quad (1)$$

where $N$ denotes the number of predictions generated, $R$ is the range of ratings in the dataset, $p_i$ is the predicted and $r_i$ is the real rating for item $i$. Low MAE reflects high accuracy of the predictions and vice versa.

### 3.2. Obfuscation in the original datasets

This experiment was aimed at examining the impact of obfuscation policies applied to the original dataset on the accuracy of the generated predictions for the entire range of ratings. For each dataset, a set of 10,000 randomly selected ratings was excluded, CF predictions were generated for these ratings, and MAE of the predictions was computed.[3] The 10,000 predictions experiment was repeated 10 times, gradually increasing the obfuscation rate

[2] The 33% and 1.5 parameters may be a basis for future experiments.

[3] Due to the large number of predicted ratings, no cross validation was performed.

**Table 2**
Properties of the original datasets.

| Dataset | Users | Items | Range | Ratings | Av. rated | Density | Average | Variance | $MAE_{np}$ |
|---|---|---|---|---|---|---|---|---|---|
| Jester | 48483 | 100 | [−10,10] | 3.52M | 72.59 | 0.726 | 0.817 | 4.400 | 0.220 |
| MovieLens | 6040 | 3952 | [1,5] | 1.00M | 165.60 | 0.042 | 3.580 | 0.935 | 0.234 |
| EachMovie | 74424 | 1649 | [0,1] | 2.81M | 37.78 | 0.023 | 0.607 | 0.223 | 0.223 |

**Table 3**
Properties of the extreme datasets.

| Dataset | Users | Items | Range | Ratings | Av. rated | Density | Average | Variance | $MAE_{np}$ |
|---|---|---|---|---|---|---|---|---|---|
| Jester | 13946 | 100 | [−10,10] | 1.01M | 72.26 | 0.723 | 0.286 | 6.111 | 0.306 |
| MovieLens | 1218 | 3952 | [1,5] | 0.18M | 144.01 | 0.036 | 3.224 | 1.116 | 0.291 |
| EachMovie | 12317 | 1649 | [0,1] | 0.49M | 39.94 | 0.024 | 0.516 | 0.379 | 0.379 |

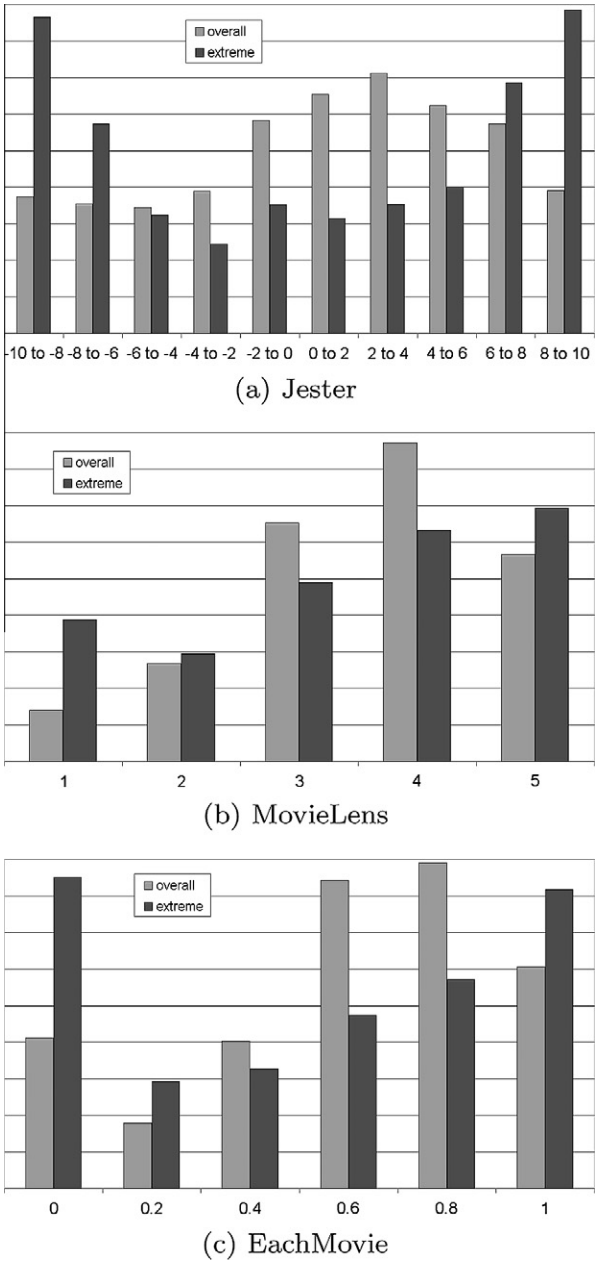


**Fig. 2.** Distribution of ratings in Jester, MovieLens, and EachMovie datasets.

from 0 (the profiles are not changed) to 0.9 (randomly selected 90% of ratings are substituted). Fig. 3 shows MAE as a function of the obfuscation rate. The horizontal axis stands for the obfuscation rate and the vertical for MAE.

The charts show that in all three datasets the impact of the *random*, *neutral*, and *distribution* policies is similar: obfuscating user profiles has a minor impact on MAE. Although MAE linearly increases with the obfuscation rate, the change in MAE is between 0.02 and 0.07, depending on the dataset. This is explained by observing that for the *random*,*neutral* and *distribution* policies, the substituted ratings are typically similar to the real ones, such that the obfuscation does not substantially modify the user profiles.

Conversely, for the *positive* and *negative* policies, the substituted ratings are extremely positive or negative. Thus, the user profiles are substantially modified and MAE increases between 0.27 and 0.35, depending on the dataset. As can be seen, the slope of the *positive* and *negative* curves is higher than of the *random*, *neutral*, and *distribution* curves. Hence, for all three datasets the impact of the *positive* and *negative* policies on the accuracy of the CF predictions is stronger than of the *random*,*neutral*, and *distribution* policies.

These results raise a question regarding the conditions for which they are valid. In other words, for which CF predictions, substituting the ratings with moderate values will substantially decrease their accuracy and for which predictions will this not happen?

### 3.3. Impact of overall obfuscation on extreme predictions

This experiment was aimed at evaluating the impact of data obfuscation on the predictions of extreme ratings. For this, the ratings in the datasets were grouped according to their value. Continuous ratings in Jester were discretized into 10 intervals: [−10,−8), [−8,− 6), [−6,−4), [−4,−2), [−2,0), [0,2), [2,4), [4,6), [6,8), and [8,10]. Discrete ratings in MovieLens and EachMovie were partitioned according to their values: 5 groups in MovieLens and 6 in EachMovie.

For each group, a set of 1,000 randomly selected ratings was excluded from the dataset, the *distribution* policy was applied, and CF predictions were generated for the excluded ratings. MAE of the predictions was computed for each group of ratings for gradually increasing from 0 to 0.9 obfuscation rate. Fig. 4 shows MAE for each group. The horizontal axis stands for the group of ratings and the vertical for MAE. For the sake of clarity, the chart shows the curves obtained for four obfuscation rates only: 0, 0.3, 0.6, and 0.9. For other obfuscation rates, the behavior of MAE was similar.

As can be seen, the impact of the obfuscation varies across the groups. For moderate ratings (central part of the chart), the MAE increase is minor. However, for extreme ratings (left and right parts of the charts), the impact of the obfuscation is stronger and MAE increases with the obfuscation rate. For high obfuscation rates, a higher MAE increase is observed for the predictions of

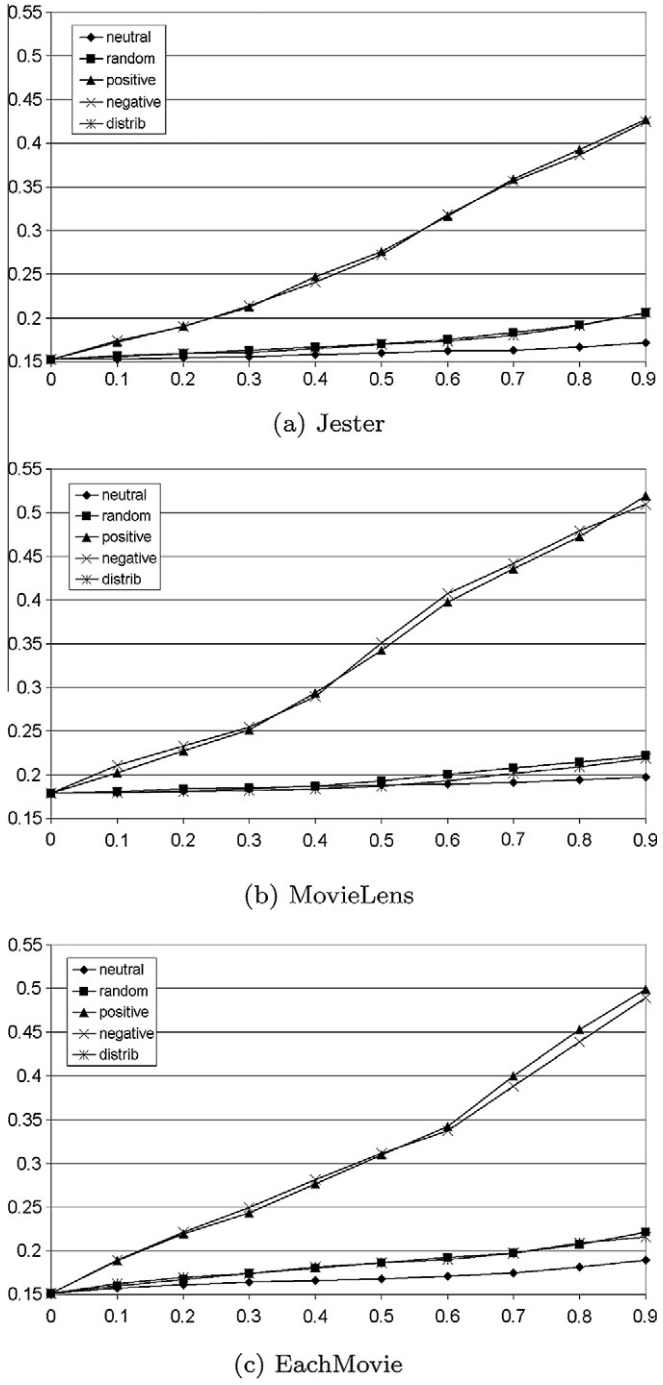


**Fig. 3.** MAE of the predictions vs. obfuscation rate in Jester, MovieLens, and EachMovie datasets.

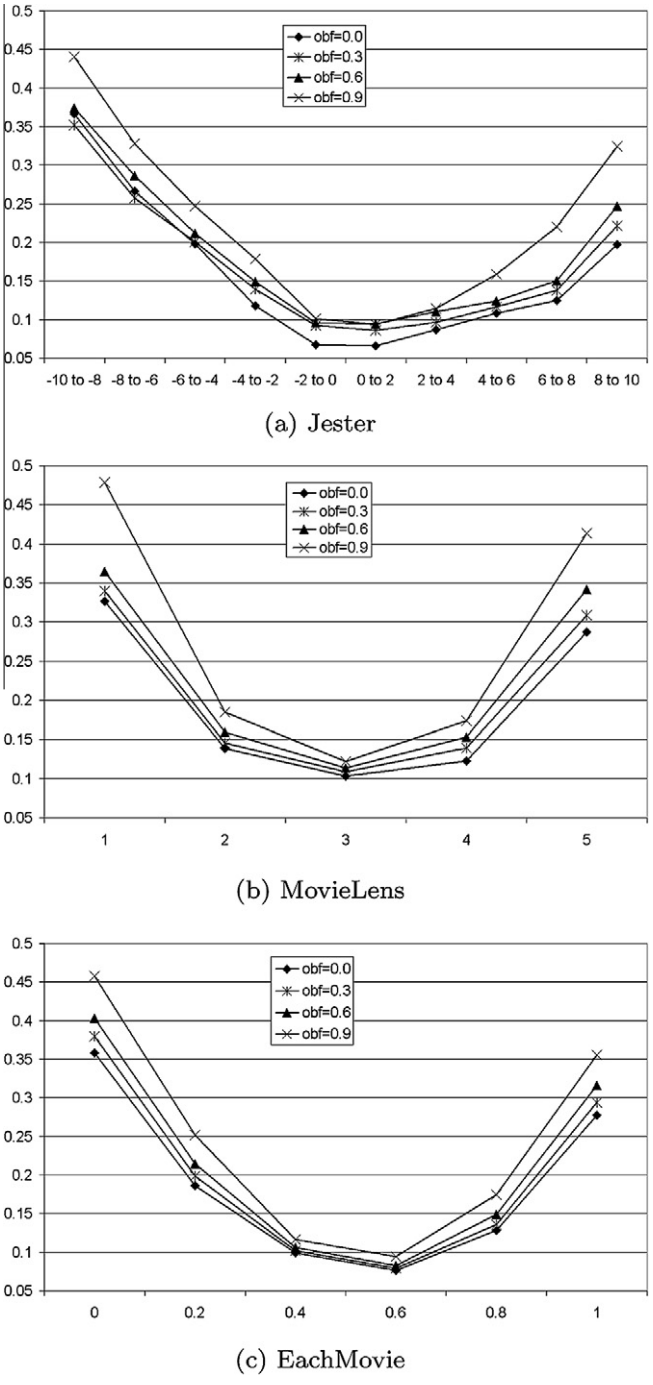


**Fig. 4.** MAE of the predictions for different groups of ratings in Jester, MovieLens, and EachMovie datasets.

extreme ratings. Thus, the accuracy of the predictions of extreme ratings is decreased by the obfuscation, while the accuracy of the predictions of moderate ratings remains unchanged.

This can be explained by considering that the *distribution* policy substitutes the real ratings with values reflecting the average and variance of ratings in the dataset. Since the average ratings of the datasets typically fall into the group of moderate ratings and their variance is low (see Table 2), applying the *distribution* policy mainly introduces moderate ratings. Hence, the obfuscation has a minor impact on the predictions of moderate ratings, as the substituted ratings are also moderate. However, it has a stronger impact on the predictions of extreme ratings, as some of the existing extreme ratings are substituted with moderate ratings and the accuracy of the predictions decreases.

### 3.4. Obfuscation in the extreme datasets

This experiment was aimed at examining the impact of the obfuscation in the extreme datasets only. For this, the obfuscation experiment similar to the experiment presented in Section 3.2 was conducted in the extreme datasets. For each extreme dataset, a set of 10,000 randomly selected ratings was excluded, CF predictions were generated for the excluded ratings, and MAE of the predictions was computed. Also this experiment was repeated 10 times,

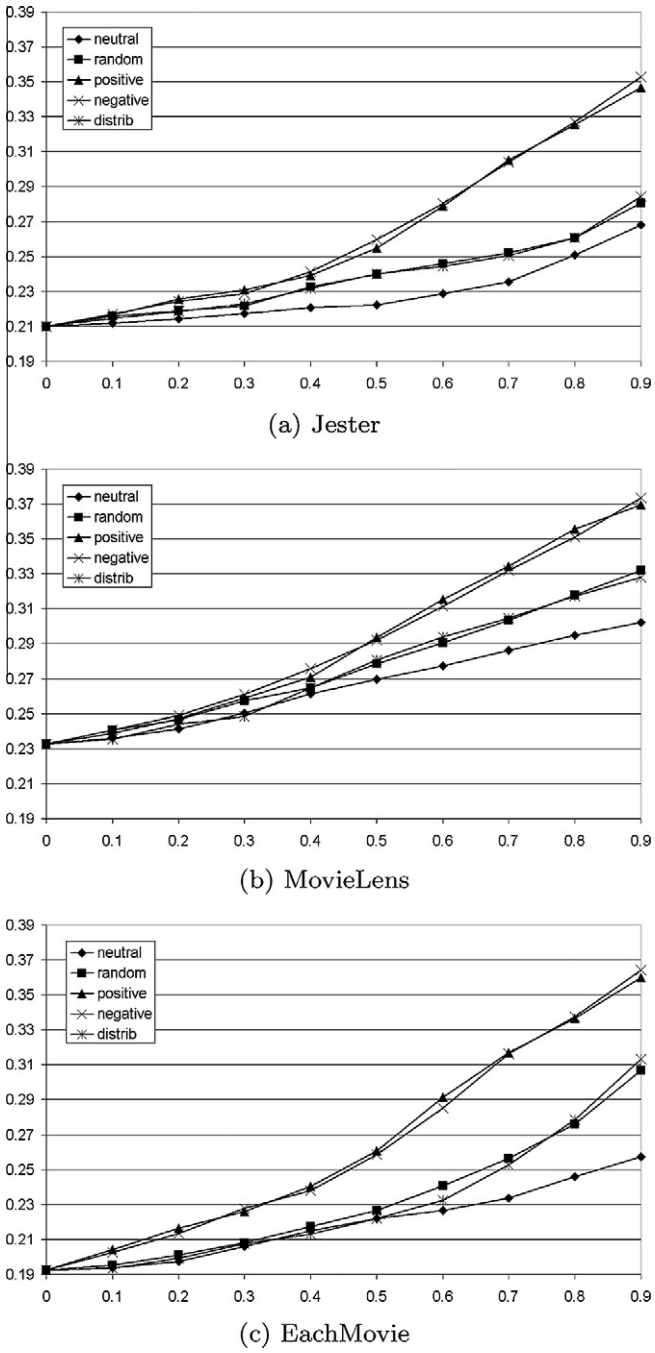


**Fig. 5.** MAE of the predictions vs. obfuscation rate in extreme Jester, MovieLens, and EachMovie datasets.

gradually increasing the obfuscation rate from 0 to 0.9. Fig. 5 shows MAE as a function of the obfuscation rate. The horizontal axis stands for the obfuscation rate and the vertical for MAE.

The results show that MAE linearly increases with the obfuscation rate for the *random*, *neutral*, and *distribution* obfuscation policies, between 0.07 and 0.12, depending on the dataset. For the *positive* and *negative* policies, the impact of the obfuscation is stronger and the change in MAE is substantially higher: between 0.14 and 0.17, depending on the dataset. Nevertheless, for the *positive* and *negative* policies, the MAE increase for the extreme datasets is lower than for the overall obfuscation experiment, where it was between 0.27 and 0.35. This is explained by observing that the extreme datasets contain more extreme ratings than the original datasets. Hence, substituting extreme ratings with extreme values does not considerably modify the data and MAE is lower than in the overall experiment.

### 3.5. Impact of localized obfuscation on extreme predictions

This experiment was aimed at evaluating the impact of a *localized* obfuscation (i.e., obfuscation of ratings having certain values) on the accuracy of the predictions. Similarly to the experiment reported in Section 3.3, the datasets were partitioned into several groups: Jester dataset was discretized into 10 groups, MovieLens partitioned into 5, and EachMovie into 6 groups.

For each dataset, a set of 10,000 randomly selected ratings from all the groups was excluded. Then the ratings in one group were obfuscated using the *distribution* policy, CF predictions for the excluded ratings were generated, and MAE of the predictions was computed. This experiment was repeated 10 times, gradually increasing the obfuscation rate from 0 to 0.9. Note that in each experiment the obfuscation was applied to ratings from a single group, i.e., only ratings having a certain value or falling within a certain range were substituted. Since the number of ratings in every group is different (see Fig. 2), MAE was normalized by dividing it by the number of ratings obfuscated in each group. Hence, the normalized MAE shows the contribution of every obfuscated rating.

Fig. 6 shows MAE for obfuscating different groups of ratings. The horizontal axis stands for the group of ratings that were obfuscated and the vertical for MAE. The chart shows the curves related to three obfuscation rates only: 0.3, 0.6, and 0.9.[4] For other obfuscation rates, the behavior of MAE is similar.

As can be seen, the impact of obfuscation varies across the groups. For moderate ratings (central part of the chart), the MAE increase is minor. However, for extreme ratings (left and right parts of the charts), the impact of the obfuscation is stronger and MAE increases with the obfuscation rate. Note that the contribution of every obfuscated rating decreases with the obfuscation rate. For low obfuscation rates, the number of obfuscated ratings is smaller than for high ones, such that the impact of every single obfuscation is stronger (although the overall impact is stronger for high obfuscation rates).

This behavior is clearly seen in Jester dataset, where MAE is an inverted bell-curve. For MovieLens, the impact of obfuscating positive ratings is weaker than of obfuscating negative ratings. This is explained by the number of positive ratings in MovieLens, which is higher than the number of negative ones (see Fig. 2). Hence, positive ratings are less extreme than negative and the impact of their obfuscation is weaker. A similar explanation is valid also for EachMovie. The abnormal behavior of the 0.2 ratings in EachMovie is explained by the number of 0.2 ratings in EachMovie, which is lower than the number of 0 and 0.4 ratings (see Fig. 2). Hence, the impact of obfuscating 0.2 ratings is stronger than of obfuscating 0 and 0.4 ratings.

In summary, the accuracy of CF predictions decreases when extreme ratings are obfuscated and remains unchanged when moderate ratings are obfuscated. This means that extreme ratings in the user profiles should be considered as the *representative* data, which are more important for generating accurate CF predictions than moderate ratings.

## 4. Attitude of users towards data obfuscation

In addition to measuring the impact of data obfuscation on the accuracy, there is a need to evaluate the users' attitude towards applying it. Some users may not be comfortable to expose the

[4] Obfuscation rate of 0 is not presented, as no ratings are modified in this case.

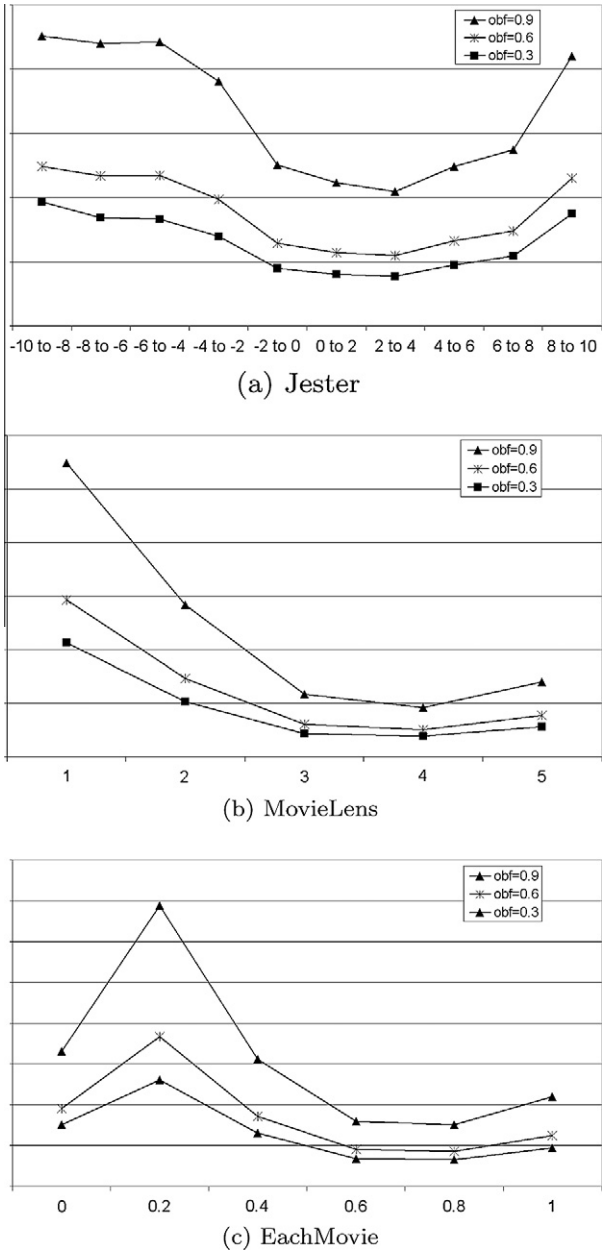


**Fig. 6.** MAE of the predictions for obfuscation of different groups of ratings in Jester, MovieLens, and EachMovie datasets.

modified ratings or may not trust less accurate recommendations. Hence, it is important to examine the attitude of users towards the privacy-personalization tradeoff introduced by the data obfuscation. This experiment is aimed at examining (1) the impact of the data obfuscation on the users' sense of privacy, (2) the perceived importance of various ratings, and (3) whether the obfuscation increases the users' willingness to expose their data.

We conducted an exploratory study, which involved 117 participants. The questions referred to a CF system using discrete ratings ranging from 1 to 5, where 1 means strongly disliking and 5 – strongly liking an item. The questions were formulated as statements and the users answered on a scale ranging from 1 to 7, where 1 means strongly disagreeing and 7 – strongly agreeing with a statement. We discretized the answers into 3 categories: 1 and 2 were treated as *disagree*, 3, 4, and 5 as *neutral*, and 6 and 7 as *agree*. Table 4 shows the average and standard deviation of answers for each question. Fig. 7 visualizes the distributions of answers (the questions will be presented shortly).

The first group of questions examined whether various ratings within one class of items are considered of different importance. Two questions were asked:

Q1: *I consider all my ratings equally sensitive, regardless of their value* (1, 2, 3, 4, *or* 5).
Q2: *I consider my ratings with extremely positive* (*equal to* 5) *and extremely negative* (*equal to* 1) *values more sensitive than other ratings* (2, 3, *or* 4).

These questions aimed at checking whether extreme ratings are considered more sensitive by the users. We defined *sensitive ratings* as "ratings the users would prefer not to make public".

We observed that answering to Q1, 47.8% of participants disagree that all the values of their ratings are equally sensitive. Furthermore, in Q2 43.0% of participants agreed that extreme ratings are more sensitive than moderate ones. Hence, the users consider their extreme ratings to be more sensitive and data obfuscation should treat the extreme ratings differently in order to enhance the users' sense of privacy.

The second group of questions examined to what extent the users are willing to expose their ratings for improving the accuracy of the generated CF predictions. Two questions were asked:

Q4: *I agree to make my average* (*equal to* 3) *ratings public, if this can improve the accuracy of the predictions provided by the system.*
Q5: *I agree to make my extremely positive* (*equal to* 5) *and extremely negative* (*equal to* 1) *ratings public, if this can improve the accuracy of the predictions provided by the system.*

Q4 examines the willingness to expose moderate ratings, while Q5 examines the willingness to expose extreme ratings.

The results showed that the users are polarized towards exposing their moderate ratings for improving the accuracy of the predictions: 34.8% of participants disagree with this, and 30.4% agree. Conversely, most users disagree to expose their extreme ratings: only 22.6% of participants agree and 53.9% disagree. Also, the averages in Table 4 validate this: the average degree of agreement to expose moderate ratings is 4.15, whereas it is 3.19 for extreme ratings. The difference was statistically significant,[5] $p < 0.05$. This strengthens our previous observation and suggests that users consider their extreme ratings to be more sensitive than moderate.

The third group of questions examined the users' appreciation of the obfuscation policies. For this, we defined the *positive*, *negative*, *neutral*, *random*, and *distribution* policies and asked a similar question for each one:

Q6: *I believe that 'positive' is a good policy to preserve my privacy.*
Q7: *I believe that 'negative' is a good policy to preserve my privacy.*
Q8: *I believe that 'neutral' is a good policy to preserve my privacy.*
Q9: *I believe that 'random' is a good policy to preserve my privacy.*
Q10: *I believe that 'distribution' is a good policy to preserve my privacy.*

The average agreement that *positive* and *negative* are good privacy-preserving policies were, respectively, 2.66 and 2.58 and most participants (56.5% for *positive* and 58.6% for *negative*) disagree that these policies are good. The average agreement for the *neutral* policy was 3.40, for *random* – 3.73, and for *distribution* – 4.009. Fewer participants disagree that these policies are good: for *neutral* – 36.7%, for *random* – 36.9%, and for *distribution* – 33.6%. Hence, *distribution* is regarded as the best privacy-preserving policy, the second best is *random*, and the third best is *neutral*. Finally, *positive* and

[5] All statistical significance tests refer to two-sample *t*-test assuming equal variances.

**Table 4**
Average answers to the study questions.

| Question | Q1 | Q2 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q13 | Q14 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | 3.212 | 4.351 | 4.148 | 3.191 | 2.657 | 2.577 | 3.404 | 3.730 | 4.009 | 4.764 | 3.694 |
| Std. dev. | 2.051 | 2.066 | 2.2 | 2.220 | 1.794 | 1.792 | 1.930 | 2.080 | 2.148 | 2.032 | 2.164 |

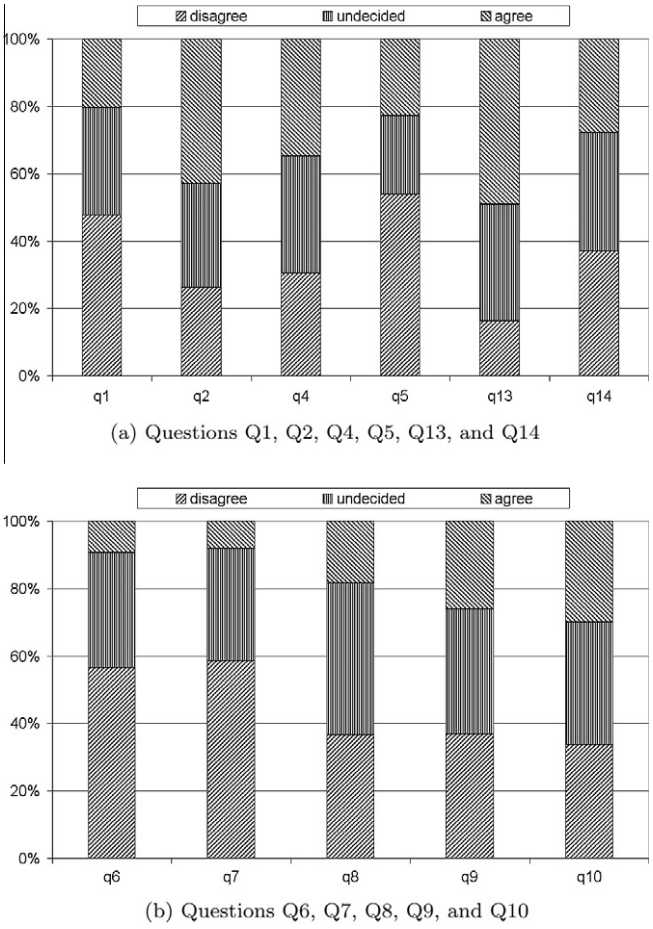


**Fig. 7.** Distribution of answers to the study questions.

*negative* are regarded as the worst privacy-preserving policies. The differences between the policies were statistically significant, $p < 0.05$, except for *positive* and *negative* policies.

These evaluations of the policies can be explained by the bias of accuracy perception rather than privacy-related considerations only. As the *positive* and *negative* policies substitute the ratings with highly dissimilar values, it is evident to users that they decrease the accuracy of the generated predictions. Hence, their evaluation of these policies is inferior to the evaluation of the *distribution*,*random*, and *neutral* policies.

The fourth group of questions was aimed at measuring whether the users' willingness to expose their ratings changes as a results of the data obfuscation. Two questions were asked:

> Q13: *I agree to make public my average (equal to* 3*) ratings, where part of them is substituted, if this can improve the accuracy of the predictions provided by the system.*
> Q14: *I agree to make public my extremely positive (equal to* 5*) and extremely negative (equal to* 1*) ratings, where part of them is substituted, if this can improve the accuracy of the predictions provided by the system.*

Q13 examines the users' willingness to expose obfuscated moderate ratings, while Q14 examines their willingness to expose obfuscated extreme ratings.

The results showed that the users increase their willingness to expose their ratings after obfuscating them. The average agreement for the moderate ratings increased from 4.15 in Q4 to 4.76 in Q13. The difference was statistically significant, $p < 0.05$. The average agreement for the extreme ratings increased from 3.19 in Q5 to 3.70 in Q14. This was also statistically significant, $p < 0.05$. Also the distribution of answers changed. Before the obfuscation, 34.8% of participants agreed to expose their moderate ratings and 22.6% – extreme ratings. After the obfuscation, these numbers increased to 49.1% and 27.8%, respectively. Hence, the willingness to expose the ratings improved as a result of the data obfuscation.

## 5. Conclusions and future research

One of the solutions to the privacy issue in CF recommender systems is to apply data obfuscation, which modifies the ratings stored in user profiles. The evaluation conducted in this work focused on the impact of data obfuscation on the accuracy of the generated predictions. The evaluation showed that users can obfuscate considerably large parts of their profiles without significantly decreasing the accuracy of the predictions.

A deeper analysis of the results yielded several interesting observations. When the experiments were conducted on the original datasets, the accuracy of the predictions was affected to a relatively minor degree. However, when only extreme ratings were considered, the accuracy of the predictions decreased. Further evaluation showed that the obfuscation of extreme ratings had a stronger impact on the accuracy of the predictions than of moderate ratings. This allowed us to conclude that extreme ratings are crucial to the accuracy of CF predictions, as they disclose the real preferences of users. This conclusion was supported by the feedback obtained from the user study: the willingness to expose extreme ratings is lower than the willingness to expose moderate ratings.

These results introduce an challenging trade-off. On one hand, the experiments show that extreme ratings are important for accurate predictions and should be exposed, while moderate ratings are less important. On the other hand, the study show that users consider their extreme ratings to be more sensitive than moderate ones and are reluctant to expose them. The combination of these two indicates that there is no simple way to optimize both the accuracy of the predictions and the users' sense of privacy. In the future, we will thoroughly investigate this issue.

Another issue raised by this work is the trade-off between the user profile obfuscation and the sparsity problem. Obfuscating the user profiles decreases the number of reliable ratings, aggravates the sparsity problem, and is expected to decrease the accuracy of the predictions. This did not happen, supposedly, due to the redundancy of ratings in the user profiles. Any user, regardless of their profile sparsity, can be classified to a certain stereotype (e.g., comedy- or drama-fan in the movie domain) based on a relatively small sample of ratings, whereas further ratings provide little new information about the user. Thus, obfuscating ratings in the user profiles decreases the redundancy, while does not increase

the data sparsity. In the future, we also plan to investigate this trade-off.

## References


Ackerman, M.S., Cranor, L.F., & Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM conference on electronic commerce* (pp. 1–8).

Agrawal, R., Kiernan, J., Srikant, R., & Xu, Y. (2004). Order-preserving encryption for numeric data. In *Proceedings of the ACM international conference on management of data* (pp. 563–574).

Androutsellis-Theotokis, S., & Spinellis, D. (2004). A survey of peer-to-peer content distribution technologies. *ACM Computing Survey, 36*(4), 335–371.

Brier, S. (1997). How to keep your privacy: Battle lines get clearer. New York Times.

Berkovsky, S., Eytani, Y., Kuflik, T., & Ricci, F. (2007). Enhancing privacy and preserving accuracy of a distributed collaborative filtering. In *Proceedings of the ACM recommender systems conference* (pp. 9–16).

Canny, J. (2002). Collaborative filtering with privacy. In *Proceedings of the IEEE symposium on security and privacy* (pp. 45–57).

Cranor, L., Reagle, J., & Ackerman, M. (1999). Beyond concern: Understanding net users attitudes about online privacy. In AT& T Labs-Research Technical Report, TR 99.4.3.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval, 4*(2), 133–151.

Herlocker, J.L., Konstan, J.A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In *Proceedings of the ACM conference on research and development in information retrieval* (pp. 230–237).

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems, 22*(1), 5–53.

Ioannidis, I., Grama, A., & Atallah, M.J. (2002). A secure protocol for computing dot-products in clustered and distributed environments. In *Proceedings of the international conference on parallel processing* (pp. 379–384).

Klösgen, W. (1995). Anonymization techniques for knowledge discovery in databases. In *Proceedings of the ACM international conference on knowledge discovery and data mining* (pp. 186–191).

Lam, S., Frankowski, D., & Riedl, J. (2006). Do you trust your recommendations? an exploration of security and privacy issues in recommender systems. In *Proceedings of the international conference on emerging trends in information and communication security* (pp. 14–29).

McJones, P. (1997). *Eachmovie collaborative filtering data set*. DEC Systems Research Center.

Miller, B. N., Konstan, J. A., & Riedl, J. (2004). Pocketlens: Toward a personal recommender system. *ACM Transactions on Information Systems, 22*(3), 437–476.

Milojicic, D.S., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B., Rollins, S., & Xu, Z. (2002). Peer-to-peer computing. In Technical Report HPL-2002-57, HP Labs.

Olsson, T. (1998). Decentralised social filtering based on trust. In *Proceedings of the AAAI recommender systems workshop* (pp. 84–88).

Parameswaran, R., & Blough, D.M. (2007). Privacy preserving collaborative filtering using data obfuscation. In *Proceedings of the IEEE international conference on granular computing* (p. 380).

Pennock, D., Horvitz, E., Lawrence, S., & Giles, C.L. (2000). Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach. In *Proceedings of the conference on uncertainty in artificial intelligence* (pp. 473–480).

Polat, H., & Du, W. (2005). Privacy-preserving collaborative filtering. *International Journal of Electronic Commerce, 9*(4), 9–35.

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *IEEE Computers, 29*(2), 38–47.

Sarwar, B.M., Karypis, G., Konstan, J.A., & Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. In *Proceedings of the ACM conference on electronic commerce* (pp. 158–167).

Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of the ACM conference on human factors in computing systems* (pp. 210–217).

Shokri, R., Pedarsani, P., Theodorakopoulos, G., & Hubaux, J.-P. (2009). Preserving privacy in collaborative filtering through distributed aggregation of offline profiles. In *Proceedings of the ACM recommender systems conference* (pp. 157–164).

Sweeney, L. (2002). K-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10*(5), 557–570.

Tveit, A. (2001). Peer-to-peer based recommendations for mobile commerce. In *Proceedings of the international workshop on mobile commerce* (pp. 26–29).

Zhang, S., Ford, J., & Makedon, F. (2006). A privacy-preserving collaborative filtering scheme with two-way communication. In *Proceedings of the ACM conference on electronic commerce* (pp. 316–323).