Enhancing Privacy and Preserving Accuracy of a Distributed Collaborative Filtering

Shlomo Berkvosky
University of Haifa, Israel
slavax@cs.haifa.ac.il

Yaniv Eytani
University of Illinois at Urbana-Champaign, USA
yeytani2@uiuc.edu

Tsvi Kuflik
University of Haifa, Israel
tsvikak@is.haifa.ac.il

Francesco Ricci
Free University of Bozen-Bolzano, Italy
fricci@unibz.it

ABSTRACT

Collaborative Filtering (CF) is a powerful technique for generating personalized predictions. CF systems are typically based on a central storage of user profiles used for generating the recommendations. However, such centralized storage introduces a severe privacy breach, since the profiles may be accessed for purposes, possibly malicious, not related to the recommendation process. Recent researches proposed to protect the privacy of CF by distributing the profiles between multiple repositories and exchange only a subset of the profile data, which is useful for the recommendation. This work investigates how a decentralized distributed storage of user profiles combined with data modification techniques may mitigate some privacy issues. Results of experimental evaluation show that parts of the user profiles can be modified without hampering the accuracy of CF predictions. The experiments also indicate which parts of the user profiles are most useful for generating accurate CF predictions, while their exposure still keeps the essential privacy of the users.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *distributed systems, user profile and alert services.*

General Terms

Algorithms, Measurement, Performance, Experimentation

Keywords

Collaborative Filtering, Recommender Systems, Privacy.

1. INTRODUCTION

Collaborative Filtering (CF) [5] is one of the most popular and widely-used personalization techniques. It generates personalized recommendations, e.g., predictions of how a user may like an item, based on the assumption that users who agreed in the past, i.e., users whose opinions correlated in the past, will also agree in the future [13]. The input for CF algorithm is a *ratings matrix* containing user profiles represented by *ratings vectors*, i.e., lists of user's ratings on a set of items. To generate a user's prediction for an item, CF initially computes the degree of similarity between the *active user*, i.e., the user whose preferences are being predicted, and all the other users. Then, CF creates a neighborhood of *K* users having the highest degree of similarity with the active user and generates a prediction for a specific item by computing a weighted average of the ratings of the other users in the neighborhood on this item.

However, personalization inherently brings with it the issue of privacy. Dealing with user profiles means that personal and possibly sensitive information about users is collected, stored and used by the recommender system. A system may violate users' privacy by misusing (e.g., selling or exposing) users' private information for their own benefits. As a result, the users that are aware and concerned about such misuse, refrain from using them to prevent potential exposure of sensitive private information [4]. Privacy hazards for recommender systems are aggravated by the fact that it is commonly believed that accurate recommendations require large amounts of personal data [11]. Thus, more complete and accurate are the user profiles, i.e., the higher is the number of ratings in the profile, the more reliable are the recommendations. Hence, there is a trade-off between the users' privacy and the accuracy of the recommendations provided to the users.

In this context, the need to protect users' privacy has triggered growing research efforts. In [3] the authors proposed basing privacy preservation on pure decentralized Peer-to-Peer (P2P) communication between the users [1]. It was suggested to form communities of users, where the overall community represents the set of users as a whole and not as individual users. Alternatively, [10] suggested preserving users' privacy on a central server by adding uncertainty to the data by applying randomized data obfuscation techniques that modify the user profiles. Hence, even if the data are exposed to untrusted parties, they will not have a reliable knowledge about the true ratings in the profiles. Current work expands and validates the idea of combining these two approaches, as initially discussed in [2]. It deals with enhancing the privacy of CF through (1) substituting the commonly used centralized CF system by a virtual P2P one, while (2) adding a degree of uncertainty to the data by modifying parts of the user profiles.

Individual users participate in the virtual P2P-based CF system in the following way. The users maintain their own profiles in form of ratings on items. Active users initiate prediction requests by exposing parts of their profiles and sending them as part of the prediction request. Other users, who actually respond to the request, expose their ratings on the requested items and similarity values with the active user, and send them to the active users, jointly with the degree of similarity between them. Note that the degree of similarity between the users was computed basing on the ratings stored by the users and part of the active user profile,

RecSys'07, October 19–02, 2007, Minneapolis, Minnesota, USA.
Copyright 2007 ACM 978-1-59593-730-8/07/0010...$5.00.

received with the prediction request. The active users collect the responses from the other users, select a subset of the most similar users as the set of nearest neighbors and aggregate the ratings of the neighbors for the prediction generation.

In this setting, the users are in full control of their personal sensitive information and they can autonomously decide when and how to expose their profiles. In particular, the users may decide that part of their profiles should be obfuscated, i.e., some noise can be added, before revealing them. As a result, the proposed approach from one hand enhances users' privacy, while from the other hand still allows them to support prediction generation initiated by other users and to participate in CF process.

In the experimental part of the paper, the accuracy of the proposed privacy-enhanced CF is evaluated using publicly available MovieLens CF dataset [5]. Initial experimental results demonstrate that there is a linear relationship between the amount of obfuscation applied to the personal ratings in the profiles and the decrease in accuracy of the recommendation prediction. These results raised a question regarding the importance of certain ratings for the accuracy of CF recommendations, i.e., about the relationship between the quality of the available data and the accuracy of the generated recommendations. Although CF is a well-studied technique, no prior works tried to understand what kind of ratings is important for the accuracy of the generated predictions. This is extremely important in the context of privacy, as users may have different concerns about the potential exposure of their data, and therefore the quantity of the user's personal data exposed to other users, must be adapted to the kind of ratings that are exposed.

For this, additional experiments aimed at analyzing the impact of data obfuscation on different types of ratings (moderate ratings with average values and extreme ratings with highly positive or highly negative values) have been conducted. The results of the experiments indicate that the accuracy of CF predictions is affected by extreme ratings stronger than by moderate ratings. Hence, the conclusion is that these parts of user profiles are the most valuable for generating accurate predictions, and for this reason they should be made available to other users. Conversely, very little knowledge about the users may be derived from their moderate ratings and, therefore, there is no need to expose these parts of the profiles.

This work also presents the results of an exploratory survey examining the users' attitude towards the privacy-preserving CF techniques illustrated in this paper. We aimed at understanding if the benefits of the proposed privacy-preserving techniques actually correlate with the users' attitude towards the techniques, i.e., if the user is convinced that the proposed techniques preserve her privacy. The results of the survey confirm that obfuscation methods having the smallest effect on the accuracy of the predictions are also preferred by the user. But they also show that the extreme ratings, which are more important for the predictions generation than the moderate ratings, are also considered by the users as more sensitive. This shows that there is no simple way to better preserve users' privacy without decreasing the accuracy of the predictions and that is difficult to optimize both the accuracy of the predictions and privacy sense of the users.

The rest of the paper is organized as follows. Section 2 discusses the privacy issues in CF and works on distributed CF. Section 3 presents the privacy-enhanced decentralized CF using user profiles obfuscation. Section 4 presents the experimental results evaluating the proposed obfuscation approach. Section 5 presents the users' survey and analyzes its results, and section 6 concludes the paper, and presents directions for future research.

2. RELATED WORKS

Centralized CF poses a severe threat to users' privacy, as personal information collected by the systems can be potentially transferred to untrusted parties. Thus, most users disagree to divulge their private information and these concerns cause some users to refrain from the benefits of recommender systems due to the privacy risks [4]. Hence, applying CF without compromising the user's privacy is one of the important and challenging issues in CF research.

This issue was tackled in prior research from several perspectives. In [10], the authors proposed a to preserve users' privacy in a centralized CF server by adding uncertainty to the data. Before transferring her profile to the server, each user obfuscated it using randomized data modification techniques. Hence, the server cannot find out the exact, but only the modified contents of the profile. Although this method changed the users' original data, experiments showed that the obfuscated data still allows generating accurate CF predictions. This approach improved users' privacy, but the users still depended on a centralized server storing the user profiles. This constituted a single point of failure, as the data could still be exposed by an attacker through a series prediction requests for various items managed by the server.

Storing user profiles distributed between several locations reduces the potential privacy breach of having all the data exposed to an attacker, as the attacker must violate security policies of all the locations, rather than of only the centralized one. Conducting CF over a distributed setting was initially proposed in [14]. This work presented a P2P architecture supporting recommendations for mobile customers represented by software agents. The agents' communication exploited an expensive routing mechanism, increasing the communication overheads. Another technique for a distributed CF eliminating the use of central servers was proposed in [8]. There, the active users create queries by sending parts of their profiles and requesting predictions for specific items. Other users autonomously decide if they are willing to respond the queries and send their information to the active users. However, no data obfuscation was applied on the data, such that the original user profiles were transferred between the users. Also, this approach was neither implemented nor evaluated.

A basic scheme for a decentralized privacy-preserving CF was proposed in [3]. According to it, individual users control their private data, while they are grouped into a community of users, representing public aggregation of their profiles. This aggregation allows personalized predictions to be computed for the members of the community or for outsiders by exposing the aggregated community data, but without exposing the data of individual users. In addition, the communication between the communities is implemented using data encryption methods. Although this approach protects overall users' privacy by abolishing a single point of failure, it puts upfront the issue of preserving the privacy of individual users, since their ratings are easier to expose than in the centralized setting. Also, the proposed approach requires a priori formation of user communities, which may become a severe limitation in nowadays dynamic environments.

In this work we combine the approaches of [3] and [10]. The users are organized in a decentralized P2P setting, where they control their profiles and participate in a distributed CF, while possible obfuscating their profile using various data perturbation techniques.

3. CF WITH DATA OBFUSCATION

This section elaborates on the prediction generation over a distributed set of users possibly obfuscating their data. It should be stressed that this work adopts pure decentralized P2P organization of users, proposed by [3]. Hence, users autonomously keep and maintain their personal profiles in pure decentralized manner. Thus, the matrix of user ratings on items, stored by centralized CF systems, is substituted by a virtual matrix, where the rows of the matrix, i.e., the ratings vectors of the users, are stored by the users in a distributed manner.

The users are connected using one of the existing P2P platforms [1]. The underlying platform guarantees connectivity of the users and allows each user to contact any of the other users connected to the system. Note that such setting does not have a single point of management or failure. Figure 1 illustrates the decentralized distribution of initially centralized ratings matrix.

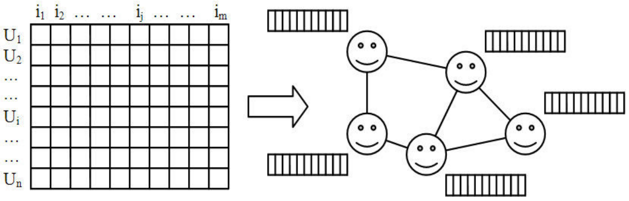

Fig. 1. Centralized vs. decentralized storage of user profiles

In this setting, users are the owners of their personal information. They directly communicate each other when during the prediction generation and independently decide about the specific ratings and parts of their profile that should be exposed to other users. The prediction generation process consists of three stages:

- The active user initiates the process through exposing her profile and broadcasting a request for a prediction for a specific item to other users. Two parameters that should be determined for this stage are:
    1. Which parts of the profile should be exposed? To better preserve the privacy of the active user, the number of ratings that are exposed should be minimized. However, decreasing the number of ratings hampers the similarity computation (as it relies on a smaller number of ratings), and therefore, hampers the accuracy of the generated predictions.
    2. To which users should the request be sent? Theoretically, the request should be sent to all the available users, since any user in the network can potentially be one of the nearest neighbors. Practically, this leads to heavy communication overheads and requires restricting the set of users to whom the request is sent.
- When the request is received, each user autonomously decides whether to respond to it. If the user decides to respond, she computes her similarity degree with the active user basing on the received parts of the active user profile. It is computed using the Cosine Similarity metric [12]. After the similarity degree is computed, this value and the user's rating on the requested item are sent to the active user. In this case two parts of the profile of the responding user are exposed: (1) the rating on the requested item, which is exposed directly, and (2) the computed similarity degree, which may allow inferring parts of the profile of the responding user.
- Upon collecting the responses, the active user builds a neighborhood of similar users for the prediction generation by selecting K users with the highest similarity degree. Finally, the active user generates a prediction for the requested item by aggregating the ratings of the neighbor users on this item as a weighted average according to their similarity degree.

To summarize the prediction generation process, it should be stressed that this form of CF preserves users' privacy (by minimizing the exposure of their profiles), while still allowing them to support predictions generation initiated by other users.

3.1 Data Obfuscation Policies

According to the above distributed CF process, the user profiles may be exposed in two cases. The first case is the profile of the active user, which is broadcasted to other users as part of the prediction request. In this case the exposure is inevitable, as the active user must expose substantial parts of her profile in order to allow a reliable similarity computation by the responding users. The second case is when the other users decide to participate in the prediction generation initiated by other user and respond to the request. The exposure of their profiles occurs when the rating on the requested item is sent to the active user for the purposes of using it at the prediction generation. Although in this case the responding users expose only a single rating from the profile, this still constitutes a privacy breach that may allow larger parts of the profiles to be exposed by the attacker through systematic malicious attacks using multiple prediction requests.

To mitigate the privacy breaches, the data in the user profiles can be obfuscated, i.e., a subset of ratings stored in the profiles can be substituted with fake values. Current work focuses on modifying the profiles of the responding users only, as modifying the profile of the active user may drastically decrease the accuracy of the similarity computation. Hence, the ratings of the responding users are substituted with fake values before computing the similarity and responding to the request. Although modifying the profiles does not prevent the attacker from collecting ratings of the responding users and reconstructing their profiles, the collected ratings will not certainly reflect the real contents of the profiles.

Several methods of modifying the data for improving privacy preservation of users' sensitive data were discussed in [7]. They include encryption, access-control policies, randomization, and anonymization. In this work, the term data *obfuscation* refers to a generalization of all the above approaches that modify the original data for the purposes of better preserving the data privacy.

In this work, three general policies for obfuscating the ratings in the user profiles are developed and experimentally compared:

- *Default obfuscation(x)* – substitute the real ratings in the user profile with a fixed predefined value *x*.
- *Uniform random obfuscation* – substitute the real ratings in the user profile with random values chosen uniformly in the range of ratings in the dataset.
- *Bell-curved random obfuscation* – substitute the real ratings in the user profile with values chosen using a bell-curve distribution reflecting the distribution of ratings in the dataset.

Supposedly, different policies have a different impact on the privacy preservation in the user profiles. For example, *default* policy substitutes the ratings with predefined values. In this case, the fake ratings are highly dissimilar from the original average ratings. Hence, there is a low probability of exposing the original user's private ratings and therefore this improves the users' privacy. Conversely, *bell-curved* policy substitutes the ratings with values reflecting the distribution of ratings in the dataset. Although in some cases the new rating may be dissimilar from the original one, overall distribution of the original and modified ratings is similar. As such, the probability of exposing ratings that are similar to the original ratings is higher, and the expected privacy improvement is lower[1].

This paper focuses on the effect of obfuscating the ratings on the accuracy of the generated predictions. The overall goal of the research is to discover specific obfuscation policies and techniques (i.e., which ratings should be substituted, to what extent, which fake values should substitute the real ratings and so forth) that facilitate a maximal preservation of users' privacy, while still allowing generation of accurate CF predictions.

3.2 Extreme Ratings and Privacy Preservation

Prior researches showed that the importance of different types of ratings for the CF process is different. For example, in [13] the authors argue that accuracy of CF is most crucial when predicting very high or very low ratings on items. This is explained by the observation that achieving high accuracy of the predictions for the best and worst items is important, while poor performance on average items is acceptable. Similarly, [9] focused on evaluating CF predictions of ratings which are *0.5* above or below the average rating in the dataset (on a scale between *0* and *5*). This is based on a similar assumption that usually the users are interested in recommendation for items she might strongly like, or indication to avoid items she might strongly dislike.

Hence, in this work the above obfuscation policies are applied on two groups of ratings: (1) obfuscating *overall ratings* – all the available ratings, and (2) obfuscating *extreme ratings* – extremely positive or extremely negative ratings only (the exact definition of extreme ratings will be given in the following section). Moreover, this work measures the effect of obfuscating the ratings in each group of ratings on the accuracy of CF predictions of two types of ratings: (1) *overall predictions* – predictions of all the available ratings, and (2) *extreme predictions* – predictions of extremely positive or extremely negative ratings.

4. EXPERIMENTAL EVALUATION

For the experimental evaluation, a decentralized environment was simulated by a multi-threaded implementation. Each user was represented by a thread and predictions were generated in the above described manner. The target user initiated the prediction generation process and broadcasted the prediction request (and her original user profile) to the other users. Upon receiving the request, each user applied the data obfuscation and modified her profile, computed the similarity degree with the target user, and returned it jointly with the rating on the requested item to the target user. Finally, the target user computed the predictions as a weighted average of the ratings of the most similar users.

The experiments were conducted on widely-used CF MovieLens [5] dataset[2]. MovieLens stores user ratings on movies, given on a discrete scale between *1* and *5*. Table 1 shows various statistical properties of the dataset: the number of users and items, the total number of ratings, the density of the dataset (i.e., the percentage of items with explicit ratings), the average and the variance of the ratings, and MAE of non-personalized predictions.

dataset	users	items	ratings	density	average	var.	$\mathbf{MAE_{np}}$
full	*6040*	*3952*	*1000209*	*0.0419*	*3.580*	*0.935*	*0.234*
extreme	*1218*	*3952*	*175400*	*0.0364*	*3.224*	*1.166*	*0.291*

Table 1. Properties of the Experimental Datasets

These parameters are shown for two datasets: *full*, containing the original MovieLens ratings, and *extreme*, containing more extreme ratings, i.e., ratings of extreme users. Extreme users were defined as users, where more than *33%* of ratings are more than *50%* farther from their average than their variance. For example, if the average rating is *3* and the variance of *0.6*, the ratings below *2.1* or above *3.9* are considered extreme. If the user profile contains *90* ratings and more than *30* are extreme, the user's ratings are extracted to the extreme dataset. Although the *33%* and *50%* thresholds are arbitrary, they filter moderate ratings and leave large dataset of extreme ratings.

To compare between the *full* and *extreme* datasets, the distribution of ratings was computed (see Table 2). As can be seen, the number of moderate ratings in the *full* dataset is significantly higher than in the *extreme* dataset, whereas for the extreme ratings the situation is opposite.

dataset	1	2	3	4	5
full	*5.62%*	*10.75%*	*26.11%*	*34.89%*	*22.63%*
extreme	*15.54%*	*11.81%*	*19.59%*	*25.32%*	*27.74%*

Table 2. Distribution of Ratings in the Datasets

In the experimental evaluation, the above mentioned three general obfuscation policies were instantiated by five specific policies:

- **Positive** – substitute the real ratings by the highest positive rating in the dataset, i.e., *5*.
- **Negative** – substitute the real ratings by the lowest negative rating in the dataset, i.e., *1*.
- **Neutral** – substitute the real ratings by the neutral rating in the dataset, i.e., an average between the maximal and minimal possible ratings, i.e., *3*.
- **Random** – substitute the real rating by a random value in the range of ratings in the dataset, i.e., from *1* to *5*.
- **Distribution** – substitute the real rating by a value reflecting the overall distribution (i.e., average and variance) of ratings in the dataset, as shown in Table 1.

Note that, *positive*, *negative* and *neutral* policies are instances of the general *default* policy, *random* policy is the instance of the general *uniform random* policy, and *distribution* policy is the general *bell-curved* policy.

Four experiments were conducted in this work. The first evaluates the impact of obfuscating overall ratings on the accuracy of

[1] This work presents a user study, which examines users' attitude towards the above obfuscation policies. In the future, we plan to quantitatively measure the privacy improvement achieved by these policies.

[2] Very similar results were obtained by conducting the experiments on Jester and EachMovie datasets. Due to the space limitations, we present only the results of MovieLens dataset.

overall predictions. The second evaluates the impact of obfuscating overall ratings on the accuracy of the predictions for ratings having various values. The third evaluates the impact of an overall obfuscation in data set of extreme ratings on the accuracy of the predictions. The fourth evaluates the impact of obfuscating ratings with different values on the accuracy of the predictions.

To evaluate the accuracy of the generated predictions, we used the well-known Mean Average Error (MAE) metric [6]:

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

where N denotes the number of the predictions, p_i is the predicted value and r_i is the real rating of the item i.

4.1 Obfuscation of Full Datasets

The first experiment was designed to examine the impact of obfuscation policies on the accuracy of the generated predictions on the full data set. For this, a set of *10,000* ratings was selected. These ratings were excluded from the dataset, their values were predicted using the distributed CF procedure described in section 3, and MAE of the predictions was computed. The *10,000* predictions were repeated *10* times, gradually increasing the obfuscation rate, i.e., increasing the amount of modified data in user profiles. The obfuscation rate increased from *0* (the original profiles are unchanged) to *0.9* (*90%* of the ratings are modified according to the applied policy). Figure 2 shows MAE values as a function of the obfuscation rate. The horizontal axis denotes the obfuscation rate and the vertical denotes MAE values.

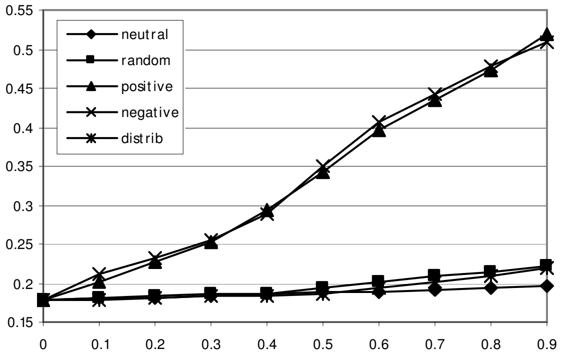

Fig. 2. MAE of the predictions vs. obfuscation rate

The graph shows that the impact of undifferentiated *random*, *neutral* and *distribution* policies is similar: MAE of the predictions increases linearly with the obfuscation rate. Although MAE increases with the obfuscation rate, the change in MAE values is between *0.018* and *0.043*, and the predictions are relatively accurate. This could be explained by the observation that *random*, *neutral* and *distribution* policies does not significantly modify the profiles of users (as the modified values are similar to the real ratings), and therefore creates a small impact on MAE. Note that for high obfuscations rates, MAE of these three policies approaches to the MAE of non-personalized predictions. Conversely, *positive* and *negative* policies modify severely the user profiles by replacing the ratings with extremely positive or negative ratings. As a result, the generated predictions are inaccurate and MAE increases rate to *0.33* and *0.34*.

This impact of the data obfuscation raises a question regarding the conditions where this observation is true. In other words, predictions of which ratings are more effected by the data obfuscation? Answering this question will allow drawing a conclusion regarding the applicability of obfuscation for the task of generating accurate CF predictions for various types of ratings.

4.2 Obfuscation vs. Extreme Predictions

To answer this question, the second experiment was aimed at evaluating the impact of data obfuscation on the predictions of various types of ratings. In this experiment, the available ratings in the dataset were partitioned to *5* groups, according to the values of the ratings: *1*, *2*, *3*, *4*, and *5*. For each group, *1,000* ratings were excluded from the dataset. *Distribution* policy was applied on the remaining ratings, and CF predictions were generated for all the excluded ratings. MAE of the predictions was computed for every group of ratings, gradually increasing the obfuscation rate from *0* to *0.9*. Figure 3 shows MAE values for various groups of ratings. The horizontal axis denotes the groups of ratings and the vertical denotes MAE values. For the sake of clarity, the graph shows only four obfuscation rates: *0*, *0.3*, *0.6* and *0.9*. For other obfuscation rates, which are not shown in this graph, the behavior of MAE curves is similar.

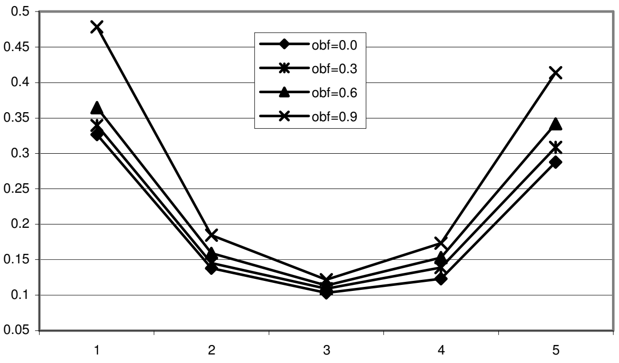

Fig. 3. MAE of the predictions for various groups of ratings

As can be seen, the impact of the obfuscation on the predictions of various types of ratings is different. For *moderate* ratings in the central part of the ratings scale, the impact of the obfuscation is minor as MAE roughly remains unchanged, regardless of the obfuscation rate. Conversely, for *extreme* (both extremely positive and negative) ratings, the impact of the obfuscation is stronger and MAE steadily increases with the obfuscation rate. Thus, the accuracy of the *extreme* ratings predictions is hampered when by the obfuscation of user profiles. Conversely, the accuracy of the *moderate* ratings predictions roughly remains unchanged regardless of the obfuscation rate.

4.3 Obfuscation of Extreme Datasets

After the special effect of extreme ratings was clarified, the third experiment examined the impact of the above obfuscation policies on the accuracy of predictions in a dataset of users with more extreme ratings. For this, the *extreme* dataset was extracted from the *full* dataset (criteria for user extremeness were described at the beginning of section 4). Then a set of *10,000* ratings was selected and excluded from the dataset. The values of these ratings were predicted and MAE of the predictions was computed. This experiment was repeated *10* times, gradually increasing the obfuscation rate from *0* to *0.9*. Figure 4 shows MAE as a function of the obfuscation rate. The horizontal axis denotes the obfuscation rate, and the vertical denotes MAE values.

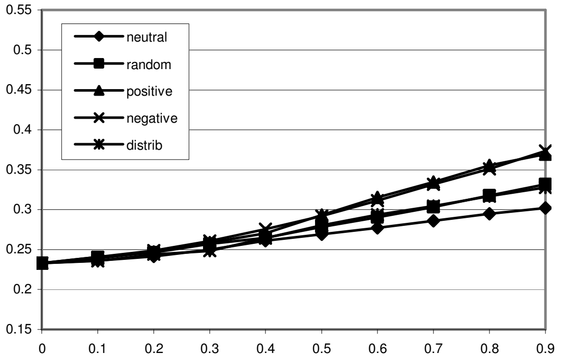

Fig. 4. MAE of the predictions vs. obfuscation rate

The experimental results show that MAE of the *extreme* dataset increases with the obfuscation rate faster than MAE of the *full* dataset. For *random*, *neutral* and *distribution* policies, the increase of MAE is between *0.069* and *0.098*. However, for *positive* and *negative* policies, the impact of data obfuscation is stronger and the increase of MAE is between *0.136%* and *0.14*. Note that in the *extreme* dataset *random*, *neutral* and *distribution* policies show a larger increase of MAE than in the *full* dataset, where it was between *0.018* and *0.043*. This can be explained by considering MAE of non-personalized predictions in extreme dataset shown in Table 2. Also in *extreme* dataset MAE of these policies approaches to the MAE of non-personalized predictions for high obfuscation rates. Since non-personalized MAE is higher than in *full* dataset, MAE values increase faster.

Conversely, for *positive* and *negative* policies, the increase of MAE is lower than in the full dataset, where it was between *0.33* and *0.34*). This is explained by the observation that most of the ratings in the *extreme* dataset are originally extreme. Hence, substituting such values with extreme values will not significantly modify the data and MAE values will be lower than in the *full* dataset experiment.

In summary, the impact of extreme ratings obfuscation on the accuracy of extreme ratings predictions is stronger than impact of overall obfuscation on the accuracy of overall predictions. In other words for a given accuracy reduction, the moderate ratings can be more extensively obfuscated compared to extreme ratings.

4.4 Extreme Obfuscation vs. Predictions

To precisely analyze the impact of obfuscation of certain ratings, in the fourth experiment we evaluated the impact of localized data obfuscation, i.e., the obfuscation of certain ratings only. For this, the available data were partitioned into *5* groups, according to the values of the ratings: *1*, *2*, *3*, *4*, and *5*, and a set of *10,000* ratings, ranging among all possible values, was selected and excluded from the dataset. Then, the values of only one group of ratings were obfuscated using *distribution* policy, the values of the excluded ratings were predicted and MAE of the predictions was computed. This experiment was repeated *10* times, gradually increasing the obfuscation rate from *0* to *0.9*. We stress that in each experiment the obfuscation was applied on the ratings of a single group of ratings only, i.e., a certain percentage of ratings with a certain value only was substituted.

It should be stressed that the obfuscation rates and MAE in this case are misleading. Since the number of ratings in every group of ratings is different (see Table 2), obfuscating a certain percentage of group ratings actually obfuscates a different number of ratings in every group and has a different impact. Hence, also MAE shows the impacts of different modified ratings. These were balanced by computing the difference between MAE for the given obfuscation rate and MAE with no obfuscation, and normalizing it by dividing the difference by the number of obfuscated ratings. This allowed us to evaluate the impact of every obfuscated rating**.**

Figure 5 shows the results of the experiments. The horizontal axis denotes the groups of ratings that were obfuscated, whereas the vertical denotes the normalized difference in MAE values. For the sake of clarity, the graph shows only four obfuscation rates: *0*, *0.3*, *0.6* and *0.9*. For other obfuscation rates, which are not shown in this graph, the behavior of MAE curves is similar.

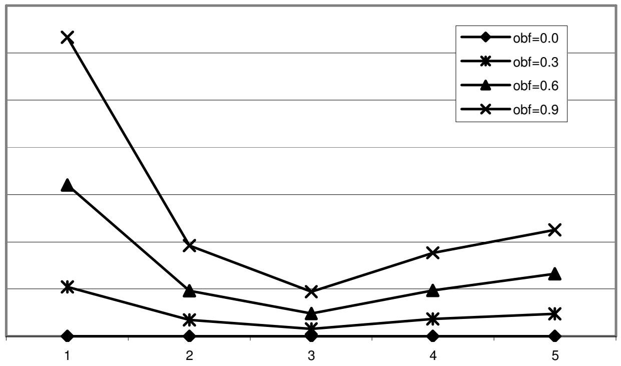

Fig. 5. Difference in MAE of the predictions for obfuscation of various ratings

As can be seen from the graph, the impact of obfuscating various types of ratings is different. Obfuscating *moderate* ratings leads to a minor increase of MAE, regardless of the obfuscation rate. Conversely, obfuscating *extreme* (both extremely positive and negative) ratings has a stronger impact on MAE. This further supports the above observations regarding the importance of extreme ratings for generation of accurate CF predictions. It must be noted that extreme negative ratings obfuscation has a larger impact on the precision. We hypothesize that it is explained by their importance in characterizing the users, as negative ratings are rare in CF data (see Table 2). In summary, the results show that the accuracy of CF predictions is hampered when the extreme ratings in user's profile are obfuscated, and the accuracy remains roughly unchanged when moderate ratings are obfuscated.

These results are particularly important for a privacy preserving CF system. They validate the trade-off between privacy and accuracy in CF, which seemed to be contradicted by the first experiment, and show that the impact of obfuscating extreme ratings is stronger than of obfuscating moderate ratings. Hence, this enables to try and adapt the obfuscation towards either the predictions accuracy or the privacy preservation.

5. ATTITUDE OF USERS SURVEY

The experiments show that the data obfuscation linearly decreases the accuracy of the generated predictions. This negative effect is compensated by an improvement of the privacy preservation of the user profiles, in the sense that less personal information is revealed. However, such an improvement may not convince the user that her privacy is actually preserved. Hence, it is important to measure the privacy gains as it is subjectively perceived by the users, i.e., to evaluate the users' attitude towards the proposed obfuscation policies and their willingness to expose their ratings before and after the obfuscation.

Privacy attitudes of users towards various types of items were studied in [4]. However, we believe that not only the items, but also the rating values within a single class of items bear different levels of importance. This is explained by the fact that the extreme ratings express a more clear preference about an item. Thus, it is important to analyze the impact of data obfuscation applied on various types of ratings on the users' sense of privacy. Also, we aim at determining if applying the data obfuscation increases users' willingness to expose their ratings.

To examine these issues, we conducted an exploratory survey with *117* subjects who responded to a request posted to several related mailing lists. The survey referred to a CF system managing numeric ratings on a scale between *1* and *5*, where *1* means disliking and *5* means liking an item. The questions were formulated as statements and the users had to express her agreement on a Likert scale ranging between *1* and *7*, where *1* means total disagreement and *7* means complete agreement with the statement. To analyze the results and neutralize personal dependencies in the answers, we partitioned the answers into three categories: answers *1-2* were evaluated as *disagree*, answers *3-5* as *neutral*, and *6-7* as *agree*. The results of the survey are presented in Table 3. It shows the distribution (in percents) and average of answers for the questions.

statement	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
agree	*20.3*	*43.0*	*34.8*	*22.6*	*9.3*	*8.1*	*18.3*	*26.2*	*29.9*	*49.1*	*27.8*
neutral	*31.9*	*30.7*	*34.8*	*23.5*	*34.3*	*33.3*	*45.0*	*36.9*	*36.4*	*34.5*	*35.2*
disagree	*47.8*	*26.3*	*30.4*	*53.9*	*56.4*	*58.6*	*36.7*	*36.9*	*33.7*	*16.4*	*37.0*
average	*3.21*	*4.35*	*4.15*	*3.19*	*2.66*	*2.58*	*3.40*	*3.73*	*4.01*	*4.76*	*3.69*

Table 3. Average Answers to the Survey Questions

The first set of questions examined if the ratings with extremely positive or extremely negative values have different importance for the users, i.e., if they are considered more sensitive by the users. The following statements were evaluated:

S1: "All my ratings are equally sensitive for me, regardless of their value (*1*, *2*, *3*, *4*, or *5*)".

S2: "My ratings with extremely positive (equal to *5*) and extremely negative (equal to *1*) values are more sensitive for me than the other ratings (*2*, *3*, or *4*)".

We observed that *47.8%* of users disagree with S1, hence claiming that not all their ratings are equally sensitive. Furthermore, in S2 *43.0%* of users agree that ratings with extremely positive or extremely negative values are more sensitive than ratings with moderate values. Hence, we can conclude that users really consider their extreme ratings as more sensitive and future privacy-enhancing algorithms should treat such ratings values differently to practically improve users' sense of privacy.

The second set of statements examined to which extent the users are willing to expose their ratings for the purpose of improving the accuracy of the generated CF predictions. The following statements were evaluated:

S3: "I agree to make my average (equal to *3*) ratings public, if this can improve the accuracy of the predictions".

S4: "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".

S3 concerns the users' willingness to expose moderate ratings, while S4 concerns their willingness to expose extreme ratings.

The results showed that the users have mixed opinions regarding exposing their moderate ratings: *30.4%* of users disagree and *34.8%* of them agree with this. However, they mostly disagree to expose their extreme ratings: only *22.6%* of users agree, while *53.9%* disagree with this. Comparing the levels of disagreement with S3 (*30.4%*) and with S4 (*53.9%*) shows that the users would share their moderate ratings more than the extreme ones. Also the average answers validate this: the average agreement for the exposure of moderate ratings is *4.15* and for the exposure of extreme ratings is *3.19* (statistically significant, *p=3.6E-09*). Intuitively, this implies that the users consider extreme rating as more sensitive and agree for their smaller exposure.

The third set of statements examined how the users evaluate various obfuscation policies described in section 4. For this, we explained the above *positive*, *negative*, *neutral*, *random* and *distribution* obfuscation policies and then we asked the users to evaluate the following statements:

S5: "*Positive* is a good policy for preserving my privacy".

S6: "*Negative* is a good policy for preserving my privacy".

S7: "*Neutral* is a good policy for preserving my privacy".

S8: "*Random* is a good policy for preserving my privacy".

S9: "*Distribution* is a good policy for preserving my privacy".

The results showed that the average levels of agreement for *positive* and *negative* obfuscation policies are *2.66* and *2.58*. Moreover, most of the users (*56.4%* for *positive* and *58.6%* for *negative*) disagree that these policies are good privacy-preserving policies. The evaluations of other three policies are slightly better. The average agreement for *neutral* policy is *3.40*, for *random* policy it is *3.73*, and for *distribution* policy it is *4.01*. Similarly, the percentage of users disagreeing that these are good privacy-preserving policies is lower. For *neutral* policy it is *36.7%*, for *random* it is *36.9%*, and for *distribution* it is *33.7%*. Hence, *distribution* policy is considered the best privacy-preserving policy, the second best is *random* policy and the third is *neutral*. Finally, *positive* and *negative* are considered the worst privacy-preserving policies (their results are almost identical). All the results are statistically significant.

We suppose that the users' evaluation of the policies is influenced by the overall evaluation of the policies and not only by privacy-related issues only. As the *positive* and *negative* policies substitute the real ratings with dissimilar values, the users may interpret that these policies introduce a strong, clearly identifiable, and wrong bias in their user profiles. Moreover, the users may prefer to keep their profile in the population average to avoid an easy identification of their (partially wrong) preferences. Hence, the evaluation of the *positive* and *negative* policies is inferior to the evaluation of the *distribution*, *random*, and *neutral* policies.

Finally, the fourth set of statements was aimed at measuring if the users' willingness to expose their ratings for improving the accuracy of the predictions has changed as a result of applying the data obfuscation. The following statements were evaluated:

S10: "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".

S11: "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".

S10 concerns the users' willingness to expose obfuscated moderate ratings, while S11 concerns their willingness to expose obfuscated extreme ratings.

The results showed that the users increased their willingness to expose both types of ratings after applying the data obfuscation. The average answer increased from *4.15* in S3 to *4.76* in S11 (statistically significant, *p=6.8E-05*) for the moderate ratings and from *3.19* in S4 to *3.69* in S12 (statistically significant, *p=9.8E-04*) for the extreme ratings. Also the distribution of answers validates our hypothesis. Prior to applying the data obfuscation, *34.8%* of users agreed to expose their moderate and *22.6%* agreed to expose their extreme ratings. After applying the obfuscation, the users' agreement increased to *49.1%* and *27.8%*, respectively.

6. CONCLUSIONS AND FUTURE WORK

This work was motivated by the need to enhance the privacy of CF personalization approach. The experimental part focused on improving the privacy preservation by applying data obfuscation and its effect on the accuracy of the generated predictions. Initial experimental results showed that the error of the predictions increases linearly with the data obfuscation rate and approaches the accuracy of non-personalized predictions. Another results showed that obfuscation of extreme ratings had a stronger effect on the accuracy of the predictions than obfuscation of moderate ratings. This allowed us to conclude that the extreme ratings are important for the accuracy of CF recommendations, as they allow identifying the real preferences of the users.

These conclusions were also validated by the opinions of the users, as shown by the results of the user survey. The survey showed that the users' willingness to expose extreme ratings is lower than their willingness to expose moderate ratings. Nevertheless, the survey that users' willingness to extreme ratings improves as a result of applying the data obfuscation.

These results introduce an important CF trade-off. From one hand, the results showed that the extreme ratings are important for generation of accurate CF predictions. Hence, these ratings should be exposed by the users to support recommendation requests initiated by other users, while the moderate ratings are less important. From the other hand, the survey showed that the users consider extreme ratings in their profiles as more sensitive and prefer not to expose them. Combination of these two conclusions highlights the trade-off between accuracy and privacy in CF indicates that there is no simple way to optimize both the accuracy of the recommendations and privacy of the users. In the future, we plan to better investigate this issue and to devise data obfuscation techniques that will be adapted both to the accuracy requirements and privacy concerns of the users. Moreover we plan to study additional obfuscation techniques, as in [16], since it has been shown that under certain conditions the randomised ratings can be fixed assuming a general consistency in the user ratings [15].

7. REFERENCES

[1] S. Androutsellis-Theotokis, D. Spinellis, "*A Survey of Peer-to-Peer Content Distribution Technologies*", in ACM Computing Survey, vol.36(4), 2004.

[2] S. Berkovsky, Y. Eytani, T. Kuflik, F. Ricci, "*Privacy-Enhanced Collaborative Filtering*", in proc. of the PEP Workshop, 2005.

[3] J. Canny, "*Collaborating Filtering with Privacy*", in proc. of the SP Symposium, 2002.

[4] L.F. Cranor, J. Reagle, M.S. Ackerman, "*Beyond Concern: Understanding Net Users' Attitudes about Online Privacy*", Technical report, AT&T Labs-Research, 1999.

[5] J.L. Herlocker, J.A. Konstan, A. Borchers, J. Riedl, "*An Algorithmic Framework for Performing Collaborative Filtering*", in proc. of the SIGIR Conference, 1999.

[6] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, "*Evaluating Collaborative Filtering Recommender Systems*", in ACM Transactions on Information Systems, vol.22(1), 2004.

[7] L. Ishitani, V. Almeida, W. Meiru, "*Masks: Bringing Anonymity and Personalization Together*", in IEEE Security and Privacy, vol. 1(3), 2003.

[8] T. Olsson, "*Decentralised Social Filtering based on Trust*", in proc. of the RS Workshop, 1998.

[9] D.M. Pennock, E. Horvitz, S. Lawrence, C.L. Giles, "*Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach*", in proc. of the UAI Conference, 2000.

[10] H. Polat, W. Du, "*Privacy-Preserving Collaborative Filtering*", in the International Journal of Electronic Commerce, vol.9(4), 2005.

[11] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "*Analysis of Recommendation Algorithms for E-Commerce*", in proc of the EC Conference, 2000.

[12] B.M. Sarwar, G. Karypis, J. Konstan, J. Riedl, "*Incremental SVD-Based Algorithms for Highly Scaleable Recommender Systems*", in proc. of the ICCIT Conference, 2002.

[13] U. Shardanand, P. Maes, "*Social Information Filtering: Algorithms for Automating 'Word of Mouth'*", in proc. of the CHI Conference, 1995.

[14] A. Tveit, "*Peer-to-Peer Based Recommendations for Mobile Commerce*", in proc. of the IWMC Workshop, 2001.

[15] N. Zhang, S. Wang, W. Zhao, "A New Scheme on Privacy-Preserving Data-Classification", in proc. of the KDD Conference, 2005.

[16] S. Zhang, J. Ford, and F. Makedon, *"A Privacy-Preserving Collaborative Filtering Scheme with Two-Way Communication"*, in proc. of the Conference on Electronic Commerce, 2006.