Data Obfuscation for Privacy-Enhanced Collaborative Filtering

Shlomo Berkovsky[1], Yaniv Eytani[2], Tsvi Kuflik[1], Francesco Ricci[3]

[1] University of Haifa, Israel, {slavax@cs,tsvikak@is}.haifa.ac.il
[2] University of Illinois at Urbana-Champaign, USA, yeytani2@uiuc.edu
[3] Free University of Bozen-Bolzano, Italy, fricci@unibz.it

Abstract

Collaborative Filtering (CF) is an attractive and reliable recommendation technique. CF is typically implemented using a centralized storage of user profiles and this is a severe privacy danger, since an attack to this central repository can endanger the quality of the recommendations and result in a leak of personal data. This work investigates how a decentralized distributed storage of user profiles combined with data obfuscation techniques can mitigate the above dangers. In an experimental evaluation we initially show that relatively large parts of the profiles can be obfuscated with a minimal increase of Mean Average Error (MAE). This contradictory result motivates further experiments where we measured the increase in prediction error in two cases: a) when a more complex prediction task is considered, i.e., a data set containing more diverse (extreme) rating values; b) when only ratings with specific values are obfuscated. The results of these experiments clarify the roles of various rating values and will help to better implement an effective obfuscation policy.

Introduction

Collaborative Filtering (CF) is probably the most popular and widely used personalization technique (Herlocker et al 1999). CF predicts user ratings for items assuming that people with observed similar ratings on a set of commonly rated items, will give similar ratings also to new unrated items (Shardanand and Maes 1995). In CF, the user profiles are represented by ratings vectors, i.e., lists of user's ratings on items. To generate a rating prediction for a given item for a target user, CF creates a neighborhood of users having a high degree of similarity with this target user, and computes a weighted average of the neighbor users' ratings on the given item.

Personalization inherently brings with it the issue of privacy. Privacy is an important challenge facing the growth of the E-Commerce services and acceptance of their various transaction models. Many services violate users' privacy for their own commercial benefits. As a result, users refrain from using them, to prevent the potential exposure of personal information (Cranor et al 1999). Privacy hazards for personalization systems are aggravated by the fact that effective personalization requires large amounts of personal data. For example, the accuracy of CF predictions is correlated with the number of similar users, number of ratings in their profiles, and the degree of their similarity (Sarwar et al 2000). Thus, the more accurate are the available user profiles, i.e., the higher is the number of ratings in the profile, the more reliable are the generated predictions. Hence, there is a trade-off between the accuracy of the personalization services provided to the users and their privacy.

The need to protect users' privacy is nowadays triggering growing research efforts in the personalization community. In (Canny 2002) the authors proposed basing privacy protection on a decentralized P2P communication between the users, hence avoiding a single point of failure and privacy breach. Alternatively, (Polat and Du 2005) suggested preserving users' privacy on a central server by adding uncertainty to the data. This was accomplished through a data obfuscation modifying the user profiles. Hence, whoever is accessing the data cannot acquire reliable knowledge about the true ratings of the users.

The combination of the above two techniques was initially proposed in (Berkovsky et al 2005). In this case, the users are in full control of the personal information stored in their profiles. Hence, they can autonomously decide when and how to expose their profiles. In particular, they may decide which parts of the profiles should be obfuscated before the exposure and may obfuscate parts of their profiles to minimize exposure of their personal data.

In this work we combine the above two ideas and mitigate privacy concerns of CF by: (1) substituting the commonly used centralized CF system with a virtual P2P one, and (2) adding some noise to the data through obfuscating parts of the user profiles. Here the added noise is considered as a way to not expose the full set of users' ratings and hence preserve their privacy. We evaluate the accuracy of the proposed privacy-enhanced CF using widely used CF MovieLens (Herlocker et al 1999) dataset and traditional MAE metric (Herlocker et al 2004). Initial experimental results surprisingly show that relatively large parts of the user profiles can be obfuscated without hampering the accuracy of the generated CF predictions.

This forced us, first, to criticize the standard MAE error measure and, second, to better analyze the impact of the obfuscation of specific parts of the user profiles. Hence, we performed two additional experiments analyzing the impact of obfuscating various types of ratings. Our results show that a) accurate CF predictions are barely affected by the obfuscation of moderate (i.e., average) ratings, b) users

with extreme ratings will be more severely penalized by obfuscation and c) obfuscation of more extreme ratings has a major impact on the prediction accuracy.

In summary, the main contribution of this paper is an architecture for privacy-preserving CF and some general experimental results showing how to better tune the obfuscation policies and evaluating tradeoffs between accuracy and privacy preservation.

The rest of the paper is organized as follows. Section 2 presents the privacy-enhanced decentralized CF. Section 3 presents the experimental results evaluating the proposed approach, and section 4 concludes the paper, and presents directions for future research.

Collaborative Filtering with Data Obfuscation

This section describes the proposed approach to generate rating predictions and recommendations over a distributed repository of users' obfuscated profiles. This work adopts a decentralized P2P organization of users, as proposed in (Canny 2002). Hence, users autonomously keep and maintain their personal profiles such that the matrix of user ratings, stored by a centralized CF system, is substituted here by a virtual matrix, whose rows, i.e., the ratings of the users, are stored directly by the users. The users are connected using a P2P communication network (Milojicic et al 2002), which (1) guarantees network connectivity regardless of joins and departures of users, and (2) facilitates communication between the connected users. Note that the evolving setting does not have a single point of management or failure. Figure 1 shows the decentralized distribution of initially centralized ratings matrix.

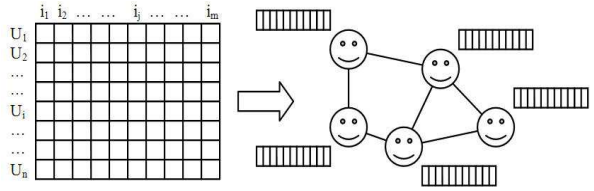

Fig. 1. Centralized vs. Decentralized Storage of the User Profiles

In this setting, users are the owners of their personal information. They directly communicate each other 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 the following stages:

- The target user initiates the process through exposing parts of her profile and broadcasting a request for a rating prediction on 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 preserve the privacy of the target user, the number of ratings that are exposed should be minimized. However, decreasing the number of ratings may hamper the similarity computation, as it will rely on a smaller number of ratings, and, therefore, hamper 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 connected user in the network can be one of the nearest neighbors of the target user. Practically, this may lead to heavy communication overheads and requires restricting the set of the users to whom the request is sent
- When the request is received, each user autonomously decides whether to respond to it. If she decides to respond, the user computes the similarity degree with the target user basing on the received part of the target user profile. When the similarity degree is computed, it is sent to the target user jointly with the rating on the requested item.
- Upon collecting the responses, the target user builds a neighborhood of similar users needed for the prediction generation. This is by selecting K users with the highest similarity degree or selecting users whose similarity degree is above a certain threshold. Then, locally aggregating the ratings of the users in the neighborhood on the requested item.

It should be stressed that this form of the CF prediction generation preserves users' privacy (by minimizing the exposure of the profiles), while allowing them to support predictions generation initiated by other users. However in this scenario the target user profile is exposed to other users and the ratings of the neighbors on the target item are exposed as well. Hence, in this distributed scenario the users may still refrain to participate in the recommendation process because of their privacy concerns.

Data Obfuscation Policies

To mitigate the above privacy breaches, we adopt data obfuscation techniques that focus on obfuscating the profiles of the responding users only, as obfuscating the target user profile may drastically decrease the accuracy of the similarity computation. Hence, parts of the responding users' profiles (i.e., certain ratings in their profiles) are substituted before computing the similarity and responding to the request. Although obfuscating the profiles does not prevent the initiator of a malicious attack to collect the ratings of the responding users, the ratings collected by an attacker will not certainly reflect the exact contents of the users' profiles.

Several data obfuscation methods were proposed for the personal data privacy preservation: encryption (Agrawal et al 2004), access-control policies (Sandhu et al 1996), data anonymization (Klosgen 1995), etc. In this work, the term *obfuscation* refers to a generalization of approaches that involve modifying the original users' data for preserving the privacy. We develop and evaluate three general policies for obfuscating the ratings in the user profiles:

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

Clearly, there is a trade-off between the amount of the substituted data in the users' profile and the accuracy of the generated predictions, as the more ratings are modified, the less accurate are the generated predictions.

Obfuscation and Privacy Preservation

As prior works have already hypothesized, the importance of various types of ratings is different (Shardanand and Maes 1995). This is particularly true in the context of privacy-preserving CF. In fact, some ratings in the user profile may be more important because they convey more personal information and could be more useful for a rating prediction. From the user's point of view, two criteria for the importance of ratings should be distinguished:

- **Content**. Certain ratings could be considered as sensitive, i.e., the users are concerned about disclosing them, because of the nature of the evaluated item. For example, such sensitive ratings are typically related to items belongings to political, sexual, and health domains.
- **Rating.** Certain ratings could be considered as sensitive simply because of their values. In fact, extreme (i.e., strongly positive or negative) ratings bring more information about the user's preferences.

This work will not measure the user perceived impact of these factors on the privacy. But we are interested in measuring how ratings of various values can be obfuscated or correctly predicted when obfuscation is used. Our ultimate goal is to link this analysis to the user perceptions and elaborate obfuscation strategies that can optimize both privacy preservation and recommendation accuracy. This paper is a first step into this direction showing the impact of obfuscation on accuracy and will be complemented by a forthcoming study on the perceived utility of the proposed obfuscation strategies from the user's point of view.

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 request to the other users. Upon receiving the request, each user 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 presented here were conducted using a commonly used CF MovieLens (Herlocker et al 1999) dataset. We note that highly similar results were obtained by conducting the same experiments on other two CF datasets: Jester and EachMovie. However, due to the space limitations, we present here the results of MovieLens only.

MovieLens contains users' ratings on movies, given on a discrete scale between *1* and *5*. Table 1 shows various parameters 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 computed by dividing the variance of ratings in the dataset by the range of ratings.

These parameters are show for two datasets: *full*, containing the full set of MovieLens ratings, and *extreme*, containing only the ratings of users defined as extreme. Extreme users were defined as users, where more than *33%* of their ratings are more than *50%* farther from the average of their ratings than their variance. For example, if the average rating is *3* and the variance of *0.6*, the ratings below *2.1* and above *3.9* are considered extreme. If the profile contains *90* ratings and more than *30* are extreme, the user's profile is extracted to the extreme dataset. Although the values of *33%* and *50%* are arbitrary (may be experimented in the future), they filter many moderate ratings and leave large enough dataset of extreme ratings.

dataset	users	items	ratings	density	average	var.	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

To analyze the extremeness of the *full* and *extreme* datasets, the distribution of ratings among their values was computed. Table 2 shows the distribution of ratings in the datasets. As can be seen, the number of moderate ratings in the *full* 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

Since the above obfuscation policies are applied by every user, she can autonomously decide (1) whether to substitute the ratings stored in the profile, (2) what percentage of ratings should be substituted (referred to in the rest of the paper as *obfuscation rate*), and (3) which ratings should be substituted. In the experiments, 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 distribution (i.e., average and variance) of the 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 overall predictions.
- The second evaluates the impact of obfuscating overall ratings on the accuracy of predictions of various types of ratings.
- The third evaluates the impact of obfuscation in data set of extreme ratings
- The fourth evaluates the impact of obfuscating various types of ratings.

To evaluate the accuracy of the generated predictions, we used the well-known Mean Average Error (MAE) metric (Herlocker et al 2004):

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

where N denotes the number of the predictions, p_i is the predicted value of the item i, and r_i is the real rating given by the user on this item. Note that low values of MAE reflect high accuracy of the predictions and vice-versa.

General Impact of Obfuscation

The first experiment was designed to examine the general impact of obfuscation policies on the accuracy of the generated predictions. For this, a set of *10,000* ratings was selected. These ratings were excluded from the dataset, their values were predicted using the above distributed CF procedure, and MAE of the predictions was computed. The *10,000* predictions experiment was 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.

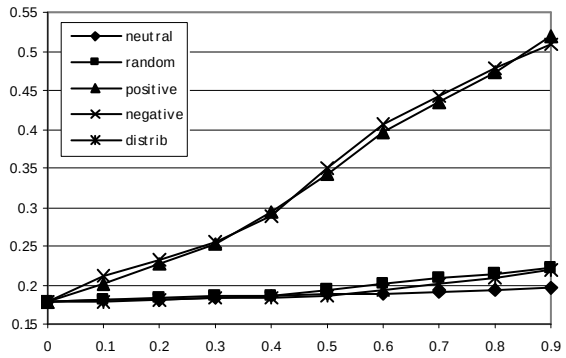

Fig. 2. MAE of the predictions vs. obfuscation rate

The graph shows that the impact of *random*, *neutral* and *distribution* policies is roughly similar: obfuscating the user profiles has a minor impact on MAE of the predictions. Although MAE slightly increases in a roughly linear manner with the obfuscation rate, the change in MAE values is between *1.8%* and *4.3%*, 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 only a small impact on MAE.

Conversely, *positive* and *negative* policies perturbate severely the user profile replacing the ratings with extremely positive or negative ratings. As a result, the generated predictions are inaccurate and MAE increases with obfuscation rate to *33%* and *34%*.

Note that for *random*, *neutral* and *distribution* policies and high obfuscation rates, the prediction accuracy is good and MAE is relatively low. In this case, MAE approaches the MAE of the non-personalized predictions.

In summary, the overall impact of an undifferentiated obfuscation on the accuracy of the predictions is minor. This raises a question regarding the conditions where this observation is true. In other words, what rating predictions are more effected by 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.

Impact of Obfuscation on the Prediction of Various Rating Values

In a second experiment we aimed at evaluating the impact of data obfuscation on the predictions of various types of ratings. In this experiment, the ratings in the MovieLens 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, 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 shows the groups of ratings and the vertical denotes MAE values. Note that for the sake of clarity, the chart shows 4 obfuscation rates only: *0*, *0.3*, *0.6* and *0.9*. For other obfuscation rates, not shown in this figure, the behavior of MAE curves is similar.

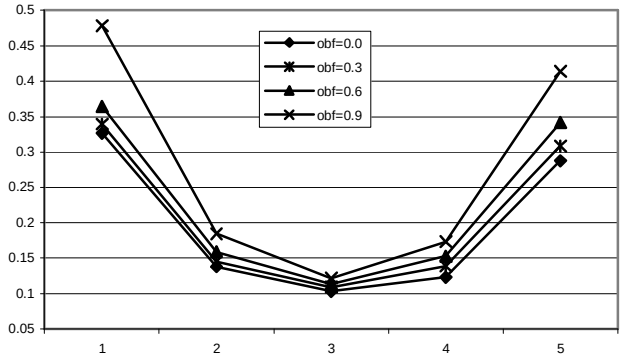

Fig. 3. MAE of the predictions for various groups of ratings

As can be clearly seen, the impact of the data obfuscation on the predictions of various types of ratings is different. For *moderate* ratings prediction, 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.

Obfuscation in a Data Set of Extreme Ratings

In a third experiment, we examined the impact of the above obfuscation policies on the accuracy of predictions in a data set composed by users with more extreme ratings. For this, the *extreme* dataset (described in Table 1) was extracted from the *full* dataset. 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 values of the obfuscation rate, and the vertical denotes MAE values.

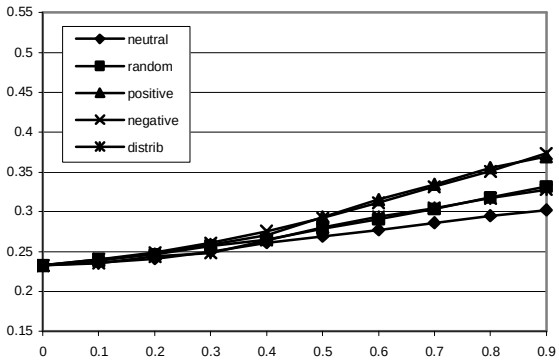

Fig. 4. MAE of the predictions vs. obfuscation rate

The experimental results clearly show that here MAE more quickly increases with the obfuscation rate. For *random*, *neutral* and *distribution* obfuscation policies, the increase of MAE is between *6.9%* and *9.8%*. However, for *positive* and *negative* policies, the negative impact of data obfuscation is stronger and the increase of MAE is between *13.6%* and *14%*. Note that in this data set *random*, *neutral* and *distribution* policies show a larger increase of MAE than in the overall dataset, where it was between *1.8%* and *4.3%*. Conversely, for *positive* and *negative* policies, the increase of MAE is lower than in the overall obfuscation experiment (was between *33%* and *34%*). This is explained by the observation that most of the ratings in the extreme ratings dataset are originally extreme. Hence, substituting such values with extreme values will not significantly modify the data in many cases and MAE values will be lower than in the overall experiment.

In summary, this shows that 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 reduction of the accuracy, the moderate ratings can be more extensively obfuscated compared to extreme ratings.

Selective Ratings Obfuscation

To precisely analyze the impact of obfuscation of certain ratings, in a fourth experiment we evaluated the impact of localized data obfuscation, i.e., the obfuscation of certain ratings only. For this, the data were partitioned to *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 a certain group of ratings were obfuscated using *distribution* policy, 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 in this case do not reliably express the amount of the obfuscated

data. 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. This was balanced by dividing the computed MAE values by the number of ratings in the respective group and, therefore comparing the relative 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 overall MAE computed on the test set (containing ratings with all possible values). Note that for the sake of clarity, the chart shows the curves related to four obfuscation rates only: *0*, *0.3*, *0.6* and *0.9*. For the other obfuscation rates not shown in this figure, the behavior of MAE curves is similar.

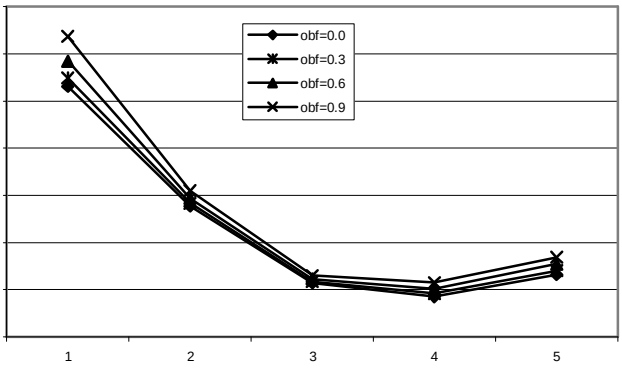

Fig. 5. MAE of the predictions for obfuscation of various ratings

As can be seen from the chart, the impact of obfuscating various types of ratings is completely different. When *moderate* ratings are obfuscated, the increase of MAE is minor, 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 observation 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. Although this may have many reasons, we conjecture that it is explained by their importance in characterizing the user preferences, as negative ratings are rare in CF data (see Table 2). In summary, this experiment shows that the accuracy of CF predictions is hampered when the *extreme* ratings in user's profile are obfuscated, especially the negative extreme ratings. Conversely, the accuracy remains roughly unchanged when moderate ratings are obfuscated.

These conclusions 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, they enable to tune the obfuscation strategy to better optimize privacy preservation and system accuracy.

Conclusions and Future Research

This work was motivated by the objective of enhancing the privacy preservation of the CF recommendation approach. We presented a distributed architecture and an obfuscation approach to preserve user privacy. We have evaluated the impact of this data obfuscation technique on the accuracy of the predictions generated by a CF based recommender system. Initial experimental results showed that relatively large parts of the profiles can be obfuscated, without hampering the accuracy of the predictions. A further analysis shown that when extreme ratings were obfuscated, the accuracy of the predictions decreased much faster with the obfuscation rate. This implies that extreme ratings are more important for generating accurate CF predictions than moderate ratings. Hence, these parts of the profiles are the most valuable for accurate predictions, and they should be preserved, while the moderate ratings are less important and can be obfuscated with a minor impact on the accuracy of the predictions. Similar results were obtained also using other CF datasets, which strengthens our conclusions.

Although the results support the conclusions regarding the importance of different parts of the profiles, they may seem controversial. One of the well-known problems of the CF is sparsity (Sarwar et al 2000), where the number of ratings in user profiles is insufficient for generating accurate predictions. One may assume that obfuscating the user profiles aggravates the sparsity. We hypothesize that this does not happen due to the user data redundancy, as the interests of a user can be determined basing on a small number of ratings. Hence, data obfuscation reduces the redundancy, but does not increase the user profiles sparsity. In the future, we plan to validate this hypothesis.

The ultimate goal of this paper is the design of a practical privacy preserving CF approach. We believe that the above results, jointly with a detailed analysis of the user perceived benefits of data obfuscation, can lead to an effective privacy-preserving recommendation technology. In fact, exploiting these results, together with a measure of the utility of various obfuscation policies presented here, allows a system designer to derive obfuscation policies, which maximize both prediction accuracy and privacy preservation. Moreover, we believe that information theory analysis of the proposed obfuscation policies can lead to identifying policies, which are better suited to support accurate prediction generation using the modified user data. In the future, we plan to investigate these issues.

References

Agrawal, R., Kiernan, J., Srikant, R., Xu, Y., 2004, “Order Preserving Encryption for Numeric Data”, in proc. of SIGMOD Conference.

Berkovsky, S., Eytani, Y., Kuflik, T., Ricci, F., 2005, “Privacy-Enhanced Collaborative Filtering”, in proc. of PEP Workshop.

Canny, J., 2002, “Collaborating Filtering with Privacy”, in proc. of SP Symposium.

Cranor, L.F., Reagle, J., Ackerman, M.S., 1999, “Beyond Concern: Understanding Net Users' Attitudes about Online Privacy”, Technical Report, AT&T Labs.

Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J., 1999, “An Algorithmic Framework for Performing Collaborative Filtering”, in proc. of SIGIR Conference.

Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T., 2004, “Evaluating Collaborative Filtering Recommender Systems”, in ACM Transactions on Information Systems, vol.22(1).

Klosgen, W., 1995, “Anonimization Techniques for Knowledge Discovery in Databases", in proc. of KDD Conference.

Milojicic, D., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B., Rollins, S., Xu, Z., 2002, “Peer-to-Peer Computing”, Technical Report, HP Labs.

Polat, H., Du, W., 2005, “Privacy-Preserving Collaborative Filtering”, in the International Journal of Electronic Commerce, vol.9(4).

Sandhu, R., Coyne, E., Feinstein, H., Youman, C., 1996, “Role-Based Access Control Models”, in IEEE Computers, vol.29(2).

Sarwar, B., Karypis, G., Konstan, J., Riedl, J., 2000, “Analysis of Recommendation Algorithms for E-Commerce”, in proc. of EC Conference.

Shardanand, U., Maes, P., 1995, “Social Information Filtering: Algorithms for Automating 'Word of Mouth'”, in proc. of CHI Conference.