# Enhancing Privacy while Preserving the Accuracy of Collaborative Filtering

**Shlomo Berkovsky[1], Tsvi Kuflik[1] and Francesco Ricci[2]**

**Abstract.** Collaborative Filtering (CF) is considered a powerful technique for generating personalized recommendations. Centralized storage of user profiles in CF systems presents a privacy breach, since the profiles are available to other users. Recent works proposed enhancing the privacy of the CF by distributing the profiles between multiple repositories. This work investigates how a decentralized distributed storage of user profiles combined with data perturbation techniques mitigates the privacy issues. Experiments, conducted on three datasets, show that relatively large parts of the profiles can be perturbed without hampering the accuracy of the CF. The experiments also allow conclusion to be drawn regarding the specific users and parts of the profiles that are valuable for generating accurate CF predictions.

## 1 INTRODUCTION

Collaborative Filtering (CF) is one of the most popular and most widely used personalization techniques. CF generates predictions of how a potential customer may like a product, based on the assumption that people with similar tastes have similar preferences for the same items [12]. In CF, the user profiles are represented by vectors of user's ratings on a set of items. To generate a prediction, CF initially creates a *neighborhood* of users having a high degree of similarity to the user whose preferences are predicted (the *active user*). Then, it generates a prediction by calculating a weighted average of the ratings of the users in the neighborhood [6].

However, personalization inherently brings with it the issue of privacy. Privacy is an important challenge facing the growth of the E-Commerce services and the acceptance of various transaction models supported by them [2]. Many services violate users' privacy for their own commercial benefits. As a result, users refrain from using them, to prevent exposure of sensitive private information [4]. 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 [11]. Thus, the more accurate are the user profiles (i.e., the higher is the number of ratings in the profile), the more reliable will be the predictions. Hence, there is a clear trade-off between the accuracy of the personalization and the privacy of users' data.

The need to protect users' privacy is nowadays triggering growing research efforts. In [3] the authors proposed basing privacy preservation on pure decentralized peer-to-peer (P2P) communication between the users. This work suggested forming user communities, where the overall community reflects the preferences of the underlying users, thus representing the set of users as whole and not as individual users. Alternatively, [10] suggested preserving users' privacy on a central server by adding uncertainty to the data. This was accomplished through using randomized perturbation techniques that modified the original user profiles. Hence, the data collector has no reliable knowledge about the true ratings of individual users. These works showed that perturbation techniques did not considerably reduce the accuracy of the generated predictions.

This paper elaborates on the idea of combining the above two techniques, as initially proposed in [1]. It suggests enhancing the privacy of the CF through substituting the commonly used centralized CF systems by virtual P2P ones, and adding a degree of uncertainty to the data through perturbing parts of the user profiles. Individual users participate in the virtual P2P-based CF system in the following way. Every user maintains his/her personal profile as a vector of rated items. Prediction is requested by an active user through exposing parts of his/her profile and sending them with a prediction request. Other users, who respond to the request, expose parts (i.e., the relevant ratings) of their profiles, and send them, jointly with locally computed degree of similarity between them and the active user. Note that in any stage, users may also perturb parts of their profiles to minimize exposure of personal data. The active user collects the responses and exploits them for neighborhood formation that leads to a local generation of the prediction.

In this setting, the users are in full control of their personal sensitive information. Hence, they can autonomously decide when and how to expose their profiles and which parts of the profiles should be perturbed before exposing them. As a result, the proposed approach enhances users' privacy, while allowing them to support prediction generation initiated by other users.

In the experimental part of the paper, the accuracy of the proposed privacy-enhanced CF is evaluated using three publicly available CF datasets: Jester [5], MovieLens [6], and EachMovie [8]. Thus, the experiments are conducted with both dense (Jester) and very sparse datasets (MovieLens and EachMovie). Results of all the datasets demonstrate that large parts of the user profiles can be perturbed without hampering the accuracy of the predictions. Also, the experimental results raise a question regarding the usefulness of the CF personalization. Although CF is considered a solid personalization technique, no prior evaluations has tried to understand which users and ratings are crucial to the accuracy of the predictions. This is important in the context of privacy, as different users have different concerns about their data, and require the quantities of the personal data exposed to be adapted accordingly. Hence, it is important to understand which parts of the user profiles (and of which users) are valuable for generating accurate predictions.

For this reason, additional experiments, aimed at analyzing the impact of data perturbation on different types of users and predictions, are conducted. The experimental results allow us to conclude that accurate CF predictions require paying a special attention to users with *extreme* preferences (either positive or negative), i.e., preferences that are significantly different from the standard preferences. Hence, these parts of the user profiles are the most valuable for generating accurate predictions, and they should be really exposed. On the other hand, usually there is no need to expose other parts of the profiles, as they provide very little knowledge about the users. It should be noted that only the user storing his/her profile can determine which parts of the profile are important for the predictions, whereas an attacker can not know it *a priori*.

[1] University of Haifa, Haifa, Israel
[2] ITC-irst, Trento, Italy

The rest of the paper is organized as follows. Section *2* discusses the privacy issues in the CF and recent works on distributed CF. Section *3* presents the basics of the privacy-enhanced decentralized CF. Section *4* presents the experimental results evaluating the proposed approach and raises a number of open questions. Section *5* analyzes the results in an attempt to answer these questions, concludes the paper, and presents directions for future research.

## 2 DISTRIBUTED CF

Collaborative filtering (CF) is a personalization technique that relies on the assumption that people who agreed in the past will also agree in the future [12]. The input for the CF algorithm is a matrix of user ratings for items, where each row represents the ratings of a single user and each column represents the ratings for a single item. CF aggregates ratings of items to recognize similarities between users, and generates a prediction for an item by weighting the ratings of similar users for the same item [6].

Centralized CF poses a threat to users' privacy, as personal information collected by service providers can be exposed and transferred to untrusted parties. Thus, most users will not agree to divulge their private information [4]. These concerns cause many users to refrain from enjoying the benefits of personalized services due to the privacy risks. Using CF without compromising user's privacy is certainly an important and challenging issue.

In [10], the authors proposed a method for preserving users' privacy on the central server by adding uncertainty to the data. Before transferring personal data to the server, each user first modifies it using randomized perturbation techniques. Therefore, the server (and also the attacker) cannot find out the exact contents of the user profiles. Although this method changes the user's original data, experiments show that the modified data still allow accurate predictions to be generated. This approach enhances users' privacy, but the users still depend on centralized, domain-specific servers. These servers constitute a single point of failure, as the data can be exposed through a series of malicious attacks.

In general, storing user profiles in several locations reduces the privacy breach of having the data exposed by the attackers, in comparison to storage on a single server. CF over a distributed setting of data repositories was initially proposed in [13]. This work presented a P2P pure decentralized architecture supporting product recommendations for mobile customers represented by software agents. The communication between the agents used an expensive routing mechanism based on network flooding that significantly increased the communication overhead. In the PocketLens project [9], five distributed architectures for the CF were implemented and compared. It was found that the performance of a P2P-based CF is close to the performance of a centralized CF.

Another technique for a distributed CF proposes eliminating the use of central servers. In this setting, an active user creates a query by sending a part of his/her profile and requesting a prediction on a specific item. Other users autonomously decide whether to respond and send information to the active user. However, this approach requires transferring the user profiles over the network, thus posing privacy breaches. Two schemes for a privacy-preserving CF were proposed in [3]. In these schemes, the users control all of their private data, while a community of users represents a public aggregation of their data, without exposing the data of individual users. The aggregation allows personalized predictions to be computed by members of the community, or by outsiders. This approach protects users' privacy in a distributed setting, but requires *a priori* formation of user groups, which becomes a severe limitation in today's dynamically evolving environments.

## 3 PRIVACY-ENHANCED CF WITH DATA PERTURBATION

This section elaborates on the prediction generation over a distributed set of users possibly perturbing their data. First, we adopt pure decentralized P2P organization of users, proposed by [3]. In this setting, users autonomously keep and maintain their personal profiles. Thus, the central matrix of user ratings on items, stored by centralized CF systems, is substituted by a virtual matrix, where every row is stored by the respective user, as shown on Figure 1.

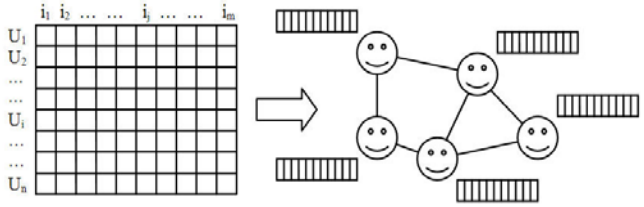


**Figure 1.** Centralized vs. decentralized storage of the user profiles

In this setting, users are the owners of their personal information. Thus, they directly communicate with each other during the prediction generation process and autonomously decide when and how to expose parts of their profiles. The predictions are generated as follows. An active user initiates the process through exposing parts of his/her profile and sending a request for a prediction on an item. When the request is received each user autonomously decides whether to respond to it. If so, he/she also exposes the relevant parts of his/her profile, (i.e., the rating for the requested item) and sends it, jointly with the locally computed degree of similarity, to the active user. Upon collecting the responses, the active user builds a neighborhood of similar users, and locally generates a prediction. This form of the CF preserves users' privacy, while allowing them to support prediction generation initiated by other users.

According to the above distributed CF, a user's profile may be exposed in two cases. The first is when a user is an active user, and he/she requests a prediction from other users. In this case the exposure is inevitable, as the active user must expose relatively large parts of his/her own profile in order to receive reliable predictions from the other users. The second case is when a user decides voluntarily to help in prediction generation initiated by other user through exposing parts of his/her profile, such that the active user can use them to build a neighborhood of similar users and generate a prediction. Although in this case the responding users have to expose relatively small parts of their profiles, this is a privacy breach that allows larger parts of the profiles to be exposed through systematic malicious attacks using multiple prediction requests.

To mitigate the privacy breaches, [10] proposes partially perturbing the data in the user profiles. We adopt the data perturbation techniques for the purposes of enhancing privacy of the P2P-based CF. For that, parts of the user profiles are modified during the process of prediction generation. Thus, in case of a malicious attack, the values collected by an attacker will not reflect the exact contents of the user's profile. Three general policies for perturbing the ratings in the user profiles were defined:

- *Uniform random obfuscation* – substitute the real ratings by random values chosen uniformly in the scope of possible ratings in the dataset.
- *Bell-curved random 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)* – real ratings in the user's profile are substituted by a predefined value *x*.

Clearly, there is a tradeoff between the amount of the perturbed data in the users' profile and the accuracy of the generated predictions, as the more data are modified, the less accurate are the generated predictions. In this sense, bell-curved obfuscation is expected to provide very accurate predictions, as the real ratings in the profiles are substituted by values that reflect the distribution of the real ratings. Nevertheless, privacy enhancement provided by it is minimal, as the exposed modified values are similar to the original ratings. Conversely, default obfuscation is expected to provide the users enhanced privacy, but very inaccurate predictions. Finally, uniform obfuscation will supposedly perform moderately in comparison to the above policies, in terms of both privacy and accuracy. The next section examines the impact of the above policies on the accuracy of the generated predictions.

## 4 EXPERIMENTAL EVALUATION

For the experimental evaluation, a pure decentralized environment was simulated by a multi-threaded implementation. Each user was represented by a thread and predictions were generated as described earlier: each thread locally computed the similarity, and returned the similarity rate and the required rating to the active thread, which computed the predictions as a weighted average of the most similar users' ratings. Hence, the process of prediction generation was performed as in a centralized CF, except for the similarity computation stage, which was done separately by each user. To measure the accuracy of the predictions, Mean Average Error (MAE) [7] values were computed by:

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

where $N$ denotes the total number of the generated predictions, $p_i$ is the prediction, and $r_i$ is the real rating for the item $i$.

To provide solid empirical evidence, the experiments were conducted using three widely used CF datasets: Jester [5], MovieLens [6] and EachMovie [8]. Table 1 summarizes the different parameters of the datasets: number of users and items in the dataset, range of ratings, total number of ratings, average number of rated items per user, density of the dataset (i.e., the percentage of items with an explicit rating), average and variance of the ratings, and the MAE of non-personalized predictions computed by dividing the variance of the ratings in the dataset by the range of ratings.

**Table 1.** Properties of the experimental datasets

| dataset | users | items | range | ratings | av.rated | density | average | var. | $MAE_{np}$ |
|---|---|---|---|---|---|---|---|---|---|
| Jester | 48483 | 100 | -10-10 | 3519449 | 72.59 | 0.7259 | 0.817 | 4.400 | 0.220 |
| ML | 6040 | 3952 | 1-5 | 1000209 | 165.60 | 0.0419 | 3.580 | 0.935 | 0.234 |
| EM | 74424 | 1649 | 0-1 | 2811718 | 37.78 | 0.0229 | 0.607 | 0.223 | 0.223 |

For the experiments, the above three perturbation policies were instantiated by five specific policies:

- *Positive* – substitute the real rating by the highest positive rating in the dataset (*10* for Jester and *5* for MovieLens and EachMovie).
- *Negative* – substitute the real rating by the lowest negative rating in the dataset (*-10* for Jester, *1* for MovieLens, and *0* for EachMovie).
- *Neutral* – substitute the real rating by the neutral rating in the dataset, i.e., an average between the maximal and minimal possible ratings (*0* for Jester, *3* for MovieLens, and 0.5 for EachMovie).
- *Random* – substitute the real rating by a random value in the range of ratings in the respective dataset (from *-10* to *10* for Jester, *1* to *5* for MovieLens, and *0* to *5* for EachMovie).
- *Distribution* – substitute the real rating by a value reflecting the distribution of the users' ratings (i.e., average and variance).

Clearly, the first three policies are instances of the *default* policy, whereas the fourth and fifth are the *uniform*, and the *bell-curved* policies discussed earlier.

The first experiment was designed to examine the impact of different perturbation policies on the accuracy of the generated predictions. In the experiment, the amount of perturbed data in user profiles was gradually increased from *0* (the profile is unchanged) to *0.9* (*90%* of the ratings in the user profiles are modified). This coefficient is referred to below as *obfuscation rate*. For each dataset, a fixed testing set of *10,000* ratings was selected, and the MAE values were computed for the possible obfuscation rates. Figure 2 shows the MAE values as a function of the obfuscation rate. The graphs refer to Jester (top), MovieLens (middle), and EachMovie (bottom) datasets. The horizontal axis denotes the obfuscation rate, whereas the vertical denotes the MAE values.

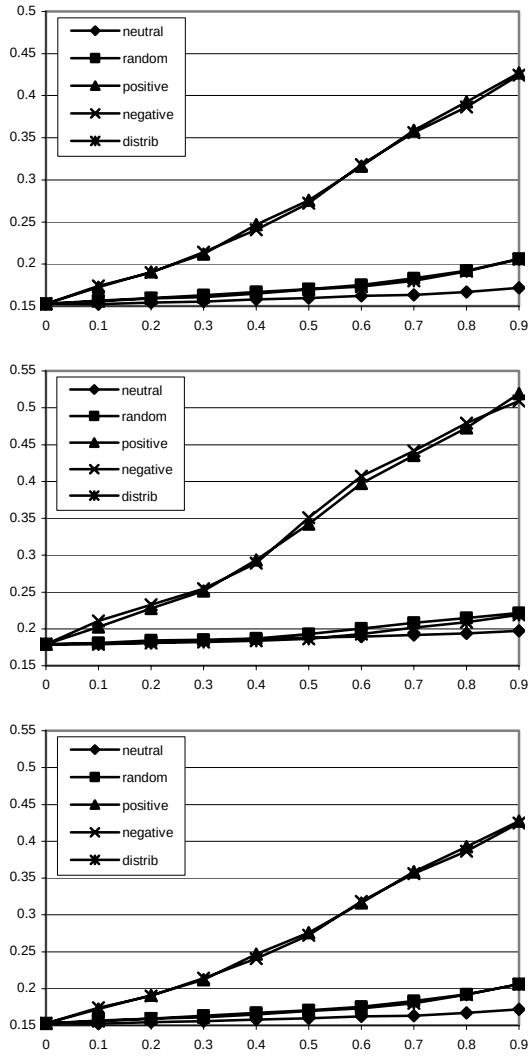


**Figure 2.** MAE of the predictions vs. obfuscation rate for Jester(top), MovieLens (middle) and EachMovie (bottom) datasets

The graphs show that the effect of *random*, *neutral* and *distribution* policies is roughly similar, as perturbing the user profiles has a minor impact on the MAE of the generated predictions. The MAE slightly increases in a roughly linear manner with the obfuscation rate; however the change is minor (between *2%* and *7%*, for differ-

ent datasets), and the prediction is still accurate. This is explained by the observation that for *random*, *neutral* and *distribution* policies, the modified values (for average users) are relatively similar to the real ratings and obfuscation does not significantly modify the user profiles. Thus, substituting the actual ratings with similar values, even for high obfuscation rates, creates only a small overall impact on the MAE computed over many users.

On the other hand, for *positive* and *negative* policies, the real ratings are substituted by highly dissimilar values. Thus, replacing the ratings with extremely positive or negative ratings does significantly modify the user profiles. As a result, the generated predictions are inaccurate and the MAE increases (maximal observed MAE values are between *27%* and *35%*, for different datasets) with obfuscation rate. The slope of the curves is significantly higher than in *random*, *neutral* and *distribution* policies. As can be clearly seen, this observation is true for all three datasets and for different levels of density in the datasets.

Note that for high obfuscation rates, despite the fact that major parts of the user profiles are perturbed, the accuracy of the predictions is good and the MAE remains low. However, comparison of CF personalized prediction MAE with non-personalized prediction MAE (taken from Table 1) yields that the MAE values are very similar for all three datasets. Thus, the MAE of CF predictions steadily converges to the MAE of non-personalized predictions with the increasing obfuscation rate. This raises a question regarding the usefulness of the CF. In other words, what are the conditions (i.e., for which users and items), where CF will generate more accurate predictions, in comparison to non-personalized predictions? This will allow us to conclude which users are valuable for CF predictions, and as such, which parts of their profiles should be exposed in order to generate accurate predictions.

To resolve this, the second experiment was designed to evaluate the impact of data perturbation on different types of predictions. For this experiment, the data of Jester[3] dataset was partitioned to *10* groups, according to the ranges of ratings: from *-10* to *-8*, to *8* to *10*. In this experiment, the *distribution* policy was exploited, and CF predictions were generated (using all the available ratings) for *1,000* ratings from each group. The MAE of the predictions was computed for gradually increasing values of the obfuscation rate. Figure 3 shows the MAE values for different groups of ratings. The horizontal axis shows the groups and the vertical denotes the MAE values. For the sake of clarity, the figure shows the curves related to four obfuscation rates only. For other obfuscation rates, the MAE values behave similarly.

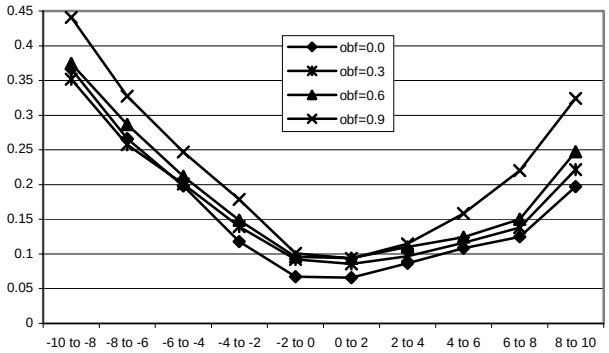


**Figure 3.** MAE of the predictions for different groups of ratings

[3] Jester dataset was selected due to the fact that Jester ratings are continuous, while MovieLens and EachMovie ratings are discrete.

As can be seen, the impact of the perturbation on different groups of ratings is different. For example, for moderate ratings, between *-2* to *2*, the impact is minor as the MAE values for all obfuscation rates are close. Conversely, for *extreme* ratings, e.g., below *-8* or above *8*, the impact of the data perturbation is clearly seen and the MAE increases with obfuscation rate. Also, for higher obfuscation rates, a larger increase in the MAE is observed for the predictions of *extreme* ratings (can be clearly seen for extremely positive ratings). Thus, the accuracy of CF predictions is hampered when extreme values in user's profile are perturbed. Practically, this means that the extreme ratings in user's profile should be considered as the user's *representative* data and they are more important for generating accurate predictions than moderate ratings.

To validate this conjecture, the first experiment was repeated on a smaller dataset of *extreme* users. The extremeness of users was defined as follows: "the user is considered as *extreme* if more than *33%* of his/her ratings are more than *50%* farther from the average of his/her ratings than his/her variance". For example, if the average rating of a user is *0.6* (on a range between *0* and *1*), and the variance is *0.2*, then the ratings below *0.3* or above *0.9* are considered as extreme ratings. If the number of ratings in user's profile is *100* and more than *33* ratings are extreme ratings, then the user is considered as extreme user. Although the selected thresholds of *33%* and *50%* are arbitrary ones (and may be a basis for future experiments), they filter out moderate users and leave large enough sets of extreme users. Table 2 summarizes the characteristics of the extreme users' datasets (columns are similar to Table 1).

**Table 2.** Properties of the extreme users' experimental datasets

| dataset | users | items | range | ratings | av.rated | density | average | var. | $MAE_{np}$ |
|---|---|---|---|---|---|---|---|---|---|
| Jester | 13946 | 100 | -10-10 | 1007700 | 72.26 | 0.7226 | 0.286 | 6.111 | 0.306 |
| ML | 1218 | 3952 | 1-5 | 175400 | 144.01 | 0.0364 | 3.224 | 1.166 | 0.291 |
| EM | 12317 | 1649 | 0-1 | 491964 | 39.94 | 0.0242 | 0.516 | 0.379 | 0.379 |

The third experiment was designed to examine the impact of the above obfuscation policies on the accuracy of the predictions for the extreme users' datasets. In the experiment, the obfuscation rate was gradually increased from *0* to *0.9*. For each dataset, a fixed testing set of *10,000* ratings was selected, and the MAE values were computed for the possible obfuscation rates. Figure 4 shows the MAE values as a function of the obfuscation rate. The graphs refer to Jester (top), MovieLens (middle), and EachMovie (bottom) datasets. The horizontal axis denotes the values of the obfuscation rate, whereas the vertical denotes the MAE values.

The experimental results clearly show that the MAE values increase with the obfuscation rate. As in the first experiment, for reasonable policies, e.g., *random*, *neutral* and *distribution*, the change in the MAE is not significant (between *7%* and *12%*). However, for *positive* and *negative* policies, the impact of data perturbation is stronger and the MAE values are significantly higher. Nevertheless, for *positive* and *negative* policies, the change in the MAE for the extreme users' dataset is lower than for the full dataset (between *14%* and *17%*). This is explained by the fact that most of the ratings in the extreme users' dataset are originally extreme. As can be seen, the MAE of personalized predictions converges to the MAE of non-personalized predictions for high obfuscation rates.

However, comparison between the extreme users' dataset and full dataset experiments yields that for extreme users the MAE values and the slope of the MAE increase are significantly higher. This allows us to conclude that extreme users (and, respectively, extreme ratings in their profiles) are important for the personalized CF prediction generation. Thus, mainly these values should be exposed by the users to support generation of accurate predictions initiated by other users.

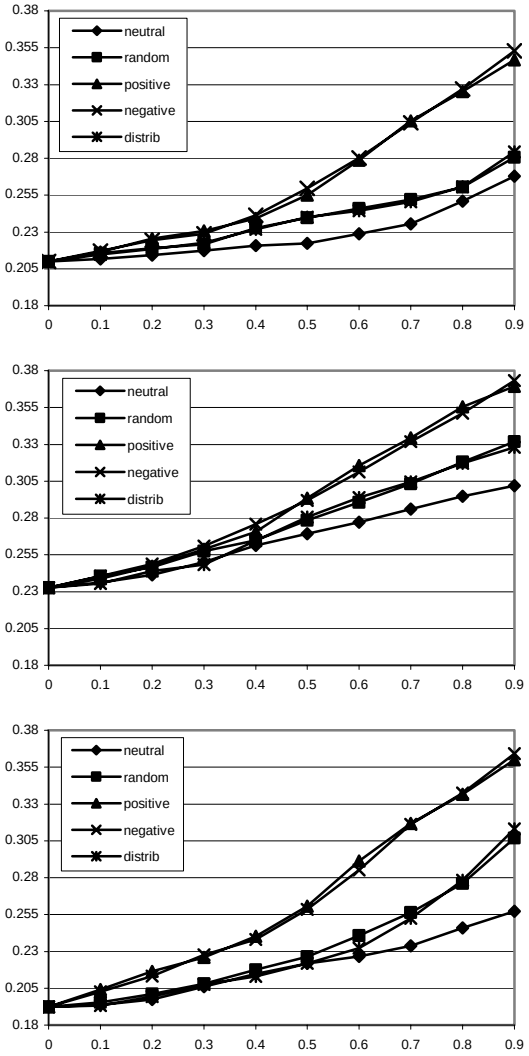


**Figure 4.** MAE of the predictions vs. obfuscation rate for Jester(top), MovieLens (middle) and EachMovie (bottom) extreme datasets

## 5 ANALYSIS, CONCLUSIONS AND FUTURE WORK

This work was motivated by enhancing the privacy of the CF personalization approach. The experimental part focused on evaluating the impact of privacy enhancement through data perturbation in distributed P2P-based environment on the accuracy of the generated predictions. Experimental results showed that relatively large parts of the user profiles can be perturbed, without hampering the accuracy of the generated predictions.

However, analysis of the results yielded interesting behavior regarding the general nature of the CF. When the experiments were conducted on a full dataset, the accuracy of the predictions was barely affected. However, when extreme ratings only were considered, the accuracy of the predictions decreased as a result of the profile obfuscations. This allowed us to conclude that the extreme ratings are the most important ratings, as they allow identifying the real preferences of the users. Hence, these parts of the user profiles are the most valuable for generating accurate predictions, and they should be exposed, while the moderate ratings are less important.

Another conclusion refers to the usefulness of the CF for personalized predictions in general. Experimental results show that the accuracy of the predictions for moderate ratings is almost not hampered by data perturbations and steadily converges to the accuracy of non-personalized predictions. Thus, heavy CF mechanism is probably not required for such predictions, and non-personalized predictions are sufficient. However, this is not true for the predictions of extreme ratings, where the accuracy is hampered by data perturbation and the MAE increases. Thus, in this case the CF is beneficial and it enhances the accuracy of the predictions.

Although the results support our conclusions regarding the importance of different parts of the profiles, they may seem contradictory. One of the well-known problems of the CF is *sparsity* [11], where the number of ratings in the user profiles is insufficient for generating accurate predictions. Perturbing parts of the profile decreases the number of reliable ratings, and aggravates the sparsity. We conjecture that this does not happen due to a data *redundancy* in the profiles. Any user, regardless of the profile sparsity, can be classified to one of the domain stereotypes (e.g., drama- or comedy-lover in movies). This classification can be done basing on a small number of extreme ratings, whereas the other ratings are redundant, i.e., they are either non-representative moderate ratings, or repeat (to some extent) the earlier ratings. Thus, perturbing randomly chosen ratings in the user profiles reduces the redundancy, while not increases the sparsity of the data in the profiles. In the future, we plan to validate this conjecture.

In the future, we also plan to discover the effect of various obfuscation policies applied to different types of ratings. In particular, we plan to validate our conclusion regarding the importance of the extreme ratings through comparing the effect of obfuscating the extreme ratings with the effect of obfuscating moderate ratings.

## REFERENCES

[1] S. Berkovsky, Y. Eytani, T. Kuflik, F. Ricci, “Privacy-Enhanced Collaborative Filtering”, in proceedings of the Workshop on Privacy-Enhanced Personalization, Edinburgh, 2005.

[2] S. Brier, “How to Keep your Privacy: Battle Lines Get Clearer”, The New York Times, 13-Jan-97.

[3] J. Canny, “Collaborating Filtering with Privacy”, in Proceedings of IEEE Conference on Security and Privacy, Oakland, 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] K. Goldberg, T. Roeder, D. Gupta, C. Perkins, “Eigentaste: A Constant Time Collaborative Filtering Algorithm”, in Information Retrieval Journal, vol.4(2), pp.133-151, 2001.

[6] J. L. Herlocker, J. A. Konstan, A. Borchers, J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering”, in proceedings of the International SIGIR Conference on Research and Development in Information Retrieval, Berkeley, 1999.

[7] 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), pp.5-53, 2004.

[8] P. McJones. “Eachmovie collaborative filtering data set”, available at http://research.compaq.com/SRC/eachmovie/, 1997.

[9] B. N. Miller, J. A. Konstan, J. Riedl, “PocketLens: Toward a Personal Recommender System”, in ACM Transactions on Information Systems, vol.22(3), pp.437-476, 2004.

[10] H. Polat, W.Du, “Privacy-Preserving Collaborative Filtering”, in the International Journal of Electronic Commerce, vol.9(4), pp.9-35, 2005.

[11] B. Sarwar, G. Karypis, J. Konstan, J.Riedl, “Analysis of Recommendation Algorithms for E-Commerce”, in proceedings of the ACM Conference on Electronic Commerce, Minneapolis, 2000.

[12] U. Shardanand, P. Maes, “Social Information Filtering: Algorithms for Automating "Word of Mouth", in proceedings of the International Conference on Human Factors in Computing Systems, Denver, 1995.

[13] A. Tveit, “Peer-to-Peer Based Recommendations for Mobile Commerce”, in proceedings of the International Workshop on Mobile Commerce, Rome, 2001.