Examining Users' Attitude towards Privacy Preserving Collaborative Filtering

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

[1] University of Haifa, Israel

[2] University of Illinois at Urbana Champaign, USA

[3] Free University of Bozen-Bolzano, Italy

[1]slavax@cs.haifa.ac.il, [1]tsvikak@is.haifa.ac.il, [2]{nikita, yeytani2}@uiuc.edu, [3]fricci@unibz.it

Abstract. Privacy hazard to Web-based information services represents an important obstacle to the growth and diffusion of the personalized services. Data obfuscation methods were proposed for enhancing the users' privacy in recommender systems based on collaborative filtering. Data obfuscation can provide statistically measurable privacy gains. However, these are measured using metrics that may not be necessarily intuitively understandable by end user, such as conditional entropy. In fact, it could happen that the users are unaware, misunderstand how their privacy is being preserved or do not feel comfortable with such methods. Thus, these may not reflect in the users' actual personal sense of privacy. In this work we provide an exploratory study to examine correlation between different data obfuscation methods and their effect on the subjective sense of privacy of users. We analyze users' opinion about the impact of data obfuscation on different types of users' rating values and generally on their sense of privacy.

1 Introduction

Web users leave identifiable tracks while surfing the Web, and there is a growing awareness of and concern about the misuse of such information [18, 22]. Many eavesdroppers on the Web violate user privacy for their own commercial benefits, and as a result, users concerned about their privacy refrain from using Web applications, just to prevent possible exposure [7]. Personalized information delivery in general, and products recommendation in particular play a major role in the development of the Web [19]. Privacy hazards for personalization systems are exacerbated by the fact that effective personalization requires large amounts of personal data. For example, consider a collaborative filtering (CF) system, a commonly used technology in the E-Commerce recommender systems [19]. In order to generate a recommendation, CF initially creates a neighborhood of users with the highest similarity to the user whose preferences are to be predicted, and it then predicts a rating for a target product (a recommendation) by averaging the ratings given by these similar users to the target item [5]. It has been shown that the accuracy of the recommendations thus generated is correlated with the number of similar users and the degree and reliability of their similarity [11] [10]. The more detailed are the user profiles and the larger their cumulative number, the more reliable will be the recommendations. Hence, there is a trade-off between the accuracy of the provided personalization and the privacy of user data.

According to a recent survey [6], most users will not agree to openly sharing their private information. However, people are not equally protective of every attribute in their data records [20, 6]. A user may not divulge the values of certain attributes at all, may not mind giving true values for others, or may be willing to share private information by giving modified values of certain attributes. Hence, in order to provide a stable dynamic infrastructure while preserving the users' privacy, a previous study [2] suggested obfuscating the user's profiles [9] by substituting part of the real values in the profiles with fake values. This setting allows users to store their personal profile locally and leaves them in control as to what personal information they would like to reveal, and when. Thus, a user requesting, for instance, similar user profiles for generating a CF recommendation, would receive only modified user profiles. From these profiles the requesting

user can learn only limited information about the true ratings of individual users. Experiments conducted with various datasets demonstrate that a relatively large part of the user profile can be obfuscated, and only a small subset of users is required to generate a recommendation with acceptable average loss in accuracy of the CF [4].

The described setting relies on the assumption that users will feel that such method does improve their actual sense of privacy and that in turn will result in their willingness to provide more personal information for the recommendation process. Further, prior CF works have highlighted that various types of CF ratings have different importance in CF. Accuracy is most crucial when predicting extreme, i.e., very high or very low, ratings on the items. This is explained by the observation that achieving high accuracy when recommending the best and worst items is most important, while poor performance on average items is acceptable [13]. Users are interested in certain predictions on items they might like or avoidance of items that might dislike, but not in precise predictions on items of which they have an average evaluation [12]. The different role played by ratings with extreme or average values is also relevant for privacy-preserving recommender systems. In fact, some ratings in the user profile are more important than the other ratings, i.e., the amount of private information encapsulated in certain ratings is higher than in other ratings.

We consider privacy enhanced personalization as a set of methods that has the characteristic of not deteriorating the prediction accuracy while at the same time use less personal data. Thus, these will leave whoever may look at the personal rating unsure about their true values. Privacy gains measure such increase in the uncertainty about original ratings found in the data. These can be by estimating by the possibility to reconstruct the distribution of the original data [2] [1]. It was previously shown [9] that data obfuscation methods provide privacy gain during the collaborative filtering process. However, such privacy metrics provide only an ordinal measure allowing comparing, on average, different methods. Further, these metrics are usually statistically oriented and thus end users may not understand how privacy is being preserved or not feel comfortable, in general, with such methods. This implies that such measurable privacy gains might not correlate with a human perception of privacy and may not reflect in the users' personal sense of privacy. Thus, it is important to examine the correlation between measurable privacy gain and its effect on the sense of privacy of users found in the system.

In addition, previous work [6] dealt mainly with the attitudes of users towards different types of items while not differentiating various ratings' values. However, we believe that not all ratings' values within one class of item (e.g., movies etc.) bear the same level of importance. This is because of their relative importance in the collaborative process as motivated before. This is also due to the fact that users intuitively express a more clear preference about an item. Thus, it is important to analyze users' opinion about the impact of privacy preserving methods to different ratings; values of ratings and the users' personal sense of privacy. In this work, we conjecture that users may want to protect ratings having extreme values (referred as extreme ratings) more carefully from being exposed. In order to examine these issues we have conducted an exploratory survey to evaluate users' opinions. Here we present our preliminary results. The main contributions of this work are:

- Assess whether users view extreme rating as being more privacy sensitive ratings (e.g., would less like to publicly share these).
- Examine to what extent users will agree to expose personal data in general, and in particular regarding to rating of different types (e.g., extreme ratings).
- Examine whether users consider different gains of their personal sense of privacy from using different types of data obfuscation policies.
- Examine whether users attitude towards sharing their personal data changes as a result of applying the obfuscation methods.

2 Data Obfuscation Policies

To provide personalization, while preserving users' privacy, [2] suggests adding uncertainty to the data by obfuscating parts of the user profiles. This reduces the amount of users' information exposed to the recommendation system, and therefore to possible malicious users getting access to the private data stored by the server. Before transferring personal data to the system, a user is

supposed to first modify her user model (products' ratings) using various perturbation techniques. Several data perturbation methods were proposed for privacy preservation of a sensitive data: encryption [14], access-control policies [15], data anonymization [16] and others. In this work, we use the term data obfuscation [17] as a generalization of all approaches that involve perturbing the data for data privacy preservation. In this context, a perturbation technique refers to the artificial modification of some of the user ratings with fake values. The rationale of this approach is that the system, and also any malicious attacker, cannot determine with certainty the exact contents of the user profiles. Although this method changes the user's original data, experiments show that it is possible to obfuscate/perturb relatively large portions of a user's profile, and still generate accurate recommendations over the modified data. The work in [4] developed and evaluates three general policies for obfuscating the ratings in the user profiles:

- *Uniform Random* obfuscation – real ratings in the user profile are substituted by random values chosen uniformly in the range of possible ratings in the dataset.
- *Curved Random* obfuscation – real ratings in the user profile are substituted by random values chosen using a bell-curve distribution with properties similar to the statistical properties of the data in the dataset (e.g., average and standard deviation of the ratings).
- *Default* obfuscation(*x*) – real ratings values in the profile are substituted by a predefined constant value *x*; Where x is highly positive, highly negative, or has a neutral value (median of the range).

Different obfuscation methods provide different mix of privacy gains (e.g., make it harder to reconstruct the original data) and loss of accuracy. For example, the *Default* obfuscation policy uses either extreme rating values or values that are close to the average rating of the dataset. Using extreme values in the obfuscation policy, has a strong negative effect on recommendation accuracy, as it substitutes the true value, which is typically close to the average, with one that is very different from the average. Moreover, these extreme ratings will clearly show some precise polarized user preference. The *Curved Random* policy reflects the actual distribution of the data and is supposed to provide the best accuracy, while preserving user privacy, since it is going to reveal a user with average preferences. Previous experiments [4] show that the obfuscated recommendation results are quite similar for different datasets with different levels of density. For instance, the effect of the random policy is an increase of the MAE (average accuracy of the predictions [21]), compared with the value obtained with no obfuscation. With high percentage of ratings perturbed with the random approach, a MAE value close to that of non-personalized recommendations is obtained. As noted before, metrics that quantify privacy gains for a given obfuscation method may not necessarily correspond with the users' sense of privacy. Hence we aim to examine how these correlate.

3 Users' extreme ratings

Prior CF works already highlighted that the importance of various types of CF ratings is different. For example, in [13] the authors argue that CF accuracy is most crucial when predicting extreme, i.e., very high or very low, ratings on the items. Intuitively, this can be explained by the observation that achieving high accuracy of the predictions on the best and worst items is most important, while poor performance on average items is acceptable. Similarly, [12] focused on evaluating CF predictions on extreme ratings, i.e., ratings which are *0.5* above or *0.5* below the average rating in the dataset (the numbers refer to a scale between *0* and *5*). This is based on a similar assumption that most of the time the users are interested in certain predictions on items they might like or denial of items that might dislike, but not in uncertain predictions on items of which they are unsure. This observation is true also in privacy-preserving issues. Some ratings in the user profile are more important than the other ratings, i.e., the amount of private information encapsulated in certain ratings is higher than in other ratings. With respect to this issue, two criteria for the importance of ratings should be distinguished: *(1) Content*: This criterion refers to the very nature of the rated items. Certain items can be considered as sensitive if the users are concerned about disclosing their opinions, i.e., their ratings, on them. For example, such sensitive items are typically related to political, sexual, religious, and health domains; *(2) Rating*: This criterion refers to the values of the ratings given by the user on the items. Clearly, extreme ratings (i.e., strongly positive

and negative evaluations) allow faster and more reliable identification of user's real preferences. Hence, disclosure and mining of private and sensitive information about the user is alleviated by presence of extreme ratings in user's profile.

In this work, we both build on the hypothesis of [13] and [12] regarding the importance of the extreme ratings during the personalization process and further correlate it with the users sense of privacy. This means, we conjectured on the importance of a ratings using the rating-based criteria and treat in a special way the ratings, whose values are extremely positive or extremely negative, rather than the ratings given on sensitive items. Hence, we aim to analyze users' opinion about the impact of privacy preserving methods to different types and values of ratings to their sense of privacy. We further would like to verify whether applying the proposed obfuscation policies will increase users' willingness to share such rating during the personalization process.

4 Examining users' personal sense of privacy

As mentioned before, measurable privacy gains may not necessarily reflect in the users' personal sense of privacy. In order to examine these issues we are currently conducting a survey to evaluate users' opinions. We defined sensitive items as follows: "*A sensitive rating is a rating you do not want to make public. For example, your ratings related to the political, sexual, religious, and health domains may be considered as sensitive*".

We obtained some preliminary results from 117 users. The rating values where supposed to be on a 1-5 scale where 1 represents disliking an item and 5 represents a highly likable item. Question replies where on a scale of 1-7 where 1 indicates strongly disagreeing and 7 represents strong agreement. Table 1 provides the average rate of agreement/disagreement for each question. Figures 1 and 2 show the distributions for the replies to each of the questions. In the figures the distribution is dived into three categories: 1-2 as disagree, 3-5 as neutral/undecided and 6-7 as agree. The survey contained 15 questions. We selected a subset of 11 questions to examine 4 issues:

First we have examined how different values of products' ratings are considered of different importance by the user within a single type of items (e.g., movies etc.). The question aims to check whether ratings with values that are extremely positive or extremely negative are conceived as more sensitive by users. This in turn implies that future algorithms should treat such ratings values differently by privacy-enhancing techniques to enhance users' personal sense of privacy.

Hypothesis: users consider extreme rating as being privacy sensitive ratings.

Q1: "All my ratings are equally sensitive for me, regardless of the value (1, 2, 3, 4, 5)."

Q2: "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, 4)."

We observed that answering to Q1 (Figure 1-left), *47.79%* of users disagree that all the values of their ratings are sensitive in the same way. Furthermore, in Q2, about *42.98%* of users strongly agree that ratings with extremely positive or extremely negative values are more sensitive than ratings with moderate values. Our results indicate that users do consider their extreme ratings as more sensitive. Thus, future privacy-enhancing algorithms should treat such ratings values differently to practically enhance users' personal sense of privacy.

The second set of questions examines whether users are willing to expose their ratings to improve predictions for other users. Q4 examines to what extent users are willing to expose their average products' ratings. Q5 is similar to Q4 but examines the issue of exposing extreme ratings.

Hypothesis: users agree to expose personal data in general, but differentiate between different types of ratings.

Q4: "I agree to make my average (equal to 3) ratings public, if this can improve the accuracy of the suggestions provided by the system."

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

The results in Figure 1-left show that users are polarized towards exposing their average ratings for the purpose of improving the accuracy of the predictions. In particular, *34.78%* of the users disagree for this, and *30.44%* of them agree. Hence, this contradicts the first part of our hypothesis that the users generally agree to expose their moderate ratings. Conversely, most of the users disagree to expose their extreme ratings: only *22.61%* of users agree to expose them, while

53.91% disagree for this. Also the average answers shown in Table 1 validate these conclusions: the average level of agreement for exposure of moderate ratings is *4.148* and for exposure of extreme ratings is *3.191*. Intuitively, these conclusions imply that users consider extreme rating as more sensitive, i.e., as more private information, and agree for a smaller exposure of extreme ratings, validating the second part of our hypothesis.

The third set of questions examines how the users evaluate the different obfuscation policies. We compare the extreme, neutral, random and overall extreme policies which are different variants of the policies described in section 2 (similar to ones defined in [4]). When describing the experimental setting we stated: "*We have designed 5 policies that can preserve your ratings' privacy. These policies aim at substituting some of your ratings with fake ratings*". Where the policies are described as follows:

- "*Positive* – substitutes the actual rating with 5, the highest possible positive rating."
- "*Negative* – substitutes the actual rating with 1, the lowest possible negative rating."
- "*Neutral* – substitutes the actual rating with 3, which is the median between the maximum and minimum possible ratings."
- "*Random* – substitutes the actual rating with a random value in the range of possible ratings (1 to 5)."
- "*Overall* – substitutes the actual rating with a random value distributed similarly to the overall distribution of all the ratings stored by the system."

Hypothesis: users view different personal sense of privacy gain from of different types of obfuscation policies.

The policies are respectively represented by questions Q6-Q10. The questions where formulated in the same way: for example, **Q6:** "I believe that the **positive policy** is a good approach for preserving my privacy."

The results show that the users' evaluations on the policies are opposite. The average levels of agreement for *positive* and *negative* obfuscation policies are, respectively, *2.657* and *2.577*. Furthermore, most of the users (*56.48%* for *positive* and *58.56%* for *negative*) disagree that these policies are good privacy-preserving mechanisms. The evaluations of the other three obfuscation policies are slightly better. The average level of agreement for the *neutral* policy is *3.404*, for the *random* policy it is *3.730*, and for the *overall* policy it is *4.009*. Similarly, the percentage of users that these policies are good privacy-preserving mechanisms is lower. For the *neutral* policy it is *36.70%*, for the *random* policy it is *36.94%*, and for the *overall* it is *33.64%*.

We hypothesize that these evaluations of the policies can be described by the effect of the general evaluation of the policies and not by privacy-related evaluation only. As the *positive* and *negative* policies substitute the real ratings with highly dissimilar fake values, they hamper the accuracy of the predictions. Hence, their general evaluations are inferior to the general evaluations of the other three policies, and the bias of the general evaluations can be seen also at privacy-related evaluations.

The forth set of questions aim to measure whether the users opinion has changed in their attitude to exposing ratings when these have been perturbed with some of the above mentioned policies. Q13 examines willingness of users to expose average ratings and Q14 similarly examines the issue regarding extreme ratings.

Hypothesis: "Users' attitude towards sharing their personal data changes as a result of applying the obfuscation policies."

Q13: "I agree to make public my average (equal to 3) ratings, where part of them is substituted, if this can improve the accuracy of the suggestions provided by the system."

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

The results clearly validate our hypothesis and show that the users increased their willingness to expose their ratings (of both types) as a result of applying the data obfuscation. The average answer regarding the moderate ratings increased from *4.148* in Q4 to *4.764* in Q13. A similar conclusion is true also for the extreme ratings as the average answer increased from *3.191* in Q5 to *3.694* in Q14. Furthermore, also the distribution of the answers validates our hypothesis. Prior to applying the data obfuscation, *34.78%* of the users agreed to expose their moderate ratings and *22.61%* agreed to expose their extreme ratings. Conversely, after applying it these numbers increased to *49.09%* and *27.78%* respectively.

Table 1. Average answers to the questions

Question	Q1	Q2	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q13	Q14
Average	3.21	4.35	4.15	3.19	2.66	2.58	3.4	3.73	4.01	4.76	3.69

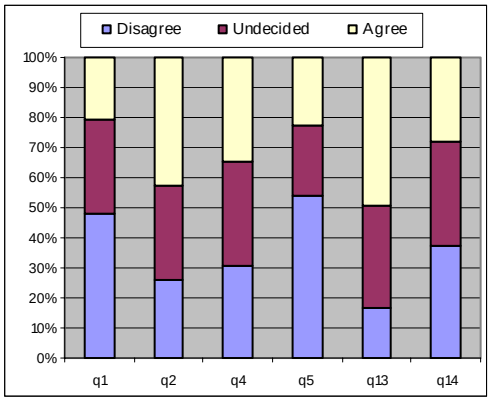

Fig. 1. Distribution of answers to Q1, Q2, Q4, Q5, Q13 and Q14

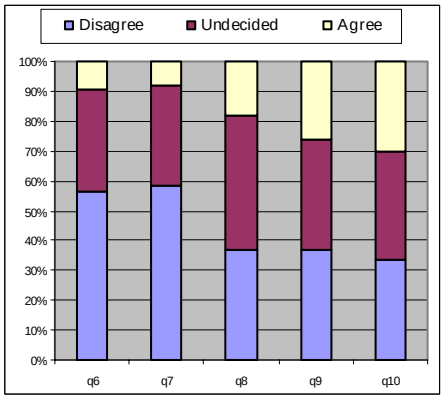

Fig. 2. Distribution of answers to Q6, Q7, Q8, Q9 and Q10

5. Discussion, conclusions and future work

We consider privacy enhanced personalization as a set of methods that has the characteristic of not deteriorating the prediction accuracy while at the same time to use less personal data. Thus, they leave unsure whoever may look at the personal rating about their true values. Privacy gains measure such uncertainty about the original ratings found in the data. It was previously shown that data obfuscation methods provide measurable privacy gains during the collaborative filtering process. However, such privacy metrics provide only an ordinal measure allowing comparing, on average, different methods. Further, these metrics are usually statistically oriented and thus end users may not understand how privacy is being preserved or generally not feel comfortable with such methods. Hence, such might not correlate with a human perception of privacy and thus may not reflect in the users' personal sense of privacy.

In addition, pervious works discuss the fact that not all ratings within one class of item (e.g., movies etc.) bears the same level of importance. Hence, it is important to analyze users' opinion about the impact of privacy preserving methods to different types of ratings to their sense of privacy. In order to examine these issues we have conducted an exploratory survey to evaluate users' opinions. This work examines the users' attitudes towards the obfuscation methods in collaborative filtering based personalization and how users would consider extreme ratings within a single type of items. Our preliminary results show that users consider extreme ratings as more sensitive and are more reluctant to expose them in the CF process. In addition users' have different attitudes towards the obfuscation methods, but in general all of them encourage users to expose their personal data. Moreover the proposed obfuscation methods seem to higher the willingness of the users to make their ratings available to the system, hence confirm the practical usability of the proposed methods.

Introducing users to the notion of privacy preserving methods when performing the survey lead them to higher willingness to share their personal data. However our current results do not allow us differentiating among the factors that lead to this inclination. Hence, future work should try to examine which are factors plays an important role for motivating users to share more of their personal data. Further, recent efforts in privacy enhanced collaborative filtering have been focusing applying it over P2P and other decentralized settings. Applying CF in such distributed setting bases on an assumption that users will feel that such methods does improve their actual sense of privacy and this in turn will result in their willingness to provide more personal information for the recommendation process. In this work we examined the former part of this assumption. We plan to examine the assumption that leaving users in control of their own profile increase their willingness to provide more information in future work. Other topic we aim to asses are how users

intuitively perceives metrics for measuring average content similarity (i.e., conditional entropy) and metrics that measure probable link-ability (i.e., anonymity sets).

Acknowledgements We thank Rajesh Kumar, Jodie P. Boyer, Ariel Gorfinkel, Dan Goldwasser and Sadek Jbara for their help during the preparation of the survey.

References

[1] D. Agrawal, C.C. Aggarwal, “On the Design and Quantification of Privacy Preserving Data Mining Algorithms, Symposium on Principles of Database Systems, 2001.

[2] R Agrawal, R. Srikant: "Privacy-Preserving Data Mining", ACM SIGMOD Int'l Conf. on Management of Data, Dallas, 2000.

[3] S. Berkovsky, Y. Eytani, T. Kuflik, F. Ricci. "Privacy-Enhanced Collaborative Filtering". In Workshop on Privacy-Enhanced Personalization (PEP), Edinburgh, UK, 2005.

[4] S. Berkovsky, Y. Eytani, T. Kuflik, F. Ricci. “Hierarchical Neighborhood Topology for Privacy Enhanced Collaborative Filtering”. In Workshop on Privacy-Enhanced Personalization (PEP), Montreal, Canada, 2006.

[5] J. Breese, D. Heckerman, C. Kadie. ”Empirical analysis of predictive algorithms for collaborative filtering.” In Uncertainty in Artificial Intelligence, Madison, WI, 1998.

[6] L.F.Cranor, J.Reagle, M.S.Ackerman, “Beyond Concern: Understanding Net Users’ Attitudes about Online Privacy”, Technical report, AT&T Labs-Research, April 1999.

[7] P.Harris, “It is Time for Rules in Wonderland”, Businessweek 20, 2000.

[8] Z. Huang, W. Du, B. Chen. “Deriving Private Information from Randomized Data.” In ACM SIGMOD Conference, , 2005, Baltimore, Maryland, USA.

[9] H. Polat, W. Du. “SVD-based Collaborative Filtering with Privacy.” In ACM Symposium on Applied Computing, Santa Fe, New Mexico,. 2005.

[10] B.M. Sarwar, G. Karypis, J.A. Konstan, J. Riedl, Analysis of recommendation algorithms for e-commerce. In *ACM Conference on Electronic Commerce*, pages 158-167, 2000.

[11] J.A. Herlocker, J. A. Konstan, A. Borchers, J. Riedl, An algorithmic framework for performing collaborative filtering. In SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 15-19, 1999, Berkeley, CA, USA, pages 230-237.

[12] D. M. Pennock, E. Horvitz, S. Lawrence, C. L. Giles, “Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach”, in proceedings of the International Conference on Uncertainty in Artificial Intelligence, Stanford, 2000.

[13] 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.

[14] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu, “Order Preserving Encryption for Numeric Data”, in proceedings of the Special Interest Group on Management of Data, Paris, 2004.

[15] R. Sandhu, E. Coyne, H. Feinstein, C. Youman, “Role-Based Access Control Models”, in IEEE Computers, vol.29(2), pp.38-47, 1996.

[16] W. Klosgen, “Anonimization Techniques for Knowledge Discovery in Databases", in proceedings of the International Conference on Knowledge and Discovery in Data Mining, Montreal, 1995.

[17] D. Bakken, R. Parameswaran, D. Blough, “Data Obfuscation: Anonymity and Desensitization of Usable Data Sets”, in IEEE Security and Privacy, vol.2(6), pp. 34-41, 2004.

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

[19] J.B. Schafer, J.A. Konstan, J. Riedl, “E-Commerce Recommendation Applications”, Journal of Data Mining and Knowledge Discovery, vol. 5 (1/2), pp. 115-152, 2001.

[20] A.F. Westin. “Freebies and Privacy: What Net Users Think”, Technical Report, Opinion Research Corporation, 1999.

[21] 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.

[22] S. Preibusch, B. Hoser, S. Gürses, Bettina Berendt, “Ubiquitous social networks' opportunities and challenges for privacy-aware user modelling”, in proceedings of the Workshop on Knowledge Discovery for Ubiquitoues User Modeling, 2007.