

Personalized Active Learning for Collaborative Filtering

Abhay S. Harpale
Carnegie Mellon University
Pittsburgh, PA 15213
aharpale@cs.cmu.edu

Yiming Yang
Carnegie Mellon University
Pittsburgh, PA 15213
yiming@cs.cmu.edu

ABSTRACT

Collaborative Filtering (CF) requires user-rated training examples for statistical inference about the preferences of new users. Active learning strategies identify the most informative set of training examples through minimum interactions with the users. Current active learning approaches in CF make an implicit and unrealistic assumption that a user can provide rating for any queried item. This paper introduces a new approach to the problem which does not make such an assumption. We *personalize* active learning for the user, and query for only those items which the user can provide rating for. We propose an extended form of Bayesian active learning and use the Aspect Model for CF to illustrate and examine the idea. A comparative evaluation of the new method and a well-established baseline method on benchmark datasets shows statistically significant improvements with our method over the performance of the baseline method that is representative for existing approaches which do not take personalization into account.

Categories and Subject Descriptors

H.4.m [Information Systems]: Miscellaneous

General Terms

Algorithms, Design, Theory, Experimentation, Human Factors

Keywords

Active Learning, Collaborative Filtering, Bayesian Active Learning, Aspect model, Personalization

1. INTRODUCTION

Collaborative Filtering (CF) [1, 2, 3, 4] has now become a popular and important technique for finding user-relevant information, e.g. movies, books and music, based on similarity of interests between groups of users. Current state-of-art machine learning algorithms cannot recommend movies,

music, and such other items, to users based on their content. CF approaches identify interest groups of users and recommend movies based on memberships of users to different interest groups. Current CF approaches can be divided into two major categories, model-based and memory-based (also known as instance-based or lazy learning). Memory-based techniques are similar to K-nearest neighbor algorithms as learning is delayed till the prediction phase. The model-based approaches make the intuitive assumption that users/items can be grouped based on their interests, and all model-based techniques apply different techniques to discover these latent or hidden *interest*-groups. Some of the popular model-based approaches are Aspect Model [2, 3], Flexible Mixture Model [4], and the Multiple Cause Vector Quantization model [5].

Active learning [6, 7, 8, 9] aims to train classifiers/models using least amount of training/labeled data, because obtaining labeled data can be a very costly or infeasible process. Active learning approaches identify instances from an unlabeled pool which will be very *beneficial* in training the model, thereby reducing the labeling cost since non-informative instances are not labeled. Active Learning approaches differ mainly in the characterization of what constitutes most beneficial training instances. Popular approaches select instances which will lead to maximum reduction in the model uncertainty/entropy [7], or which will lead to maximum reduction in prediction error [6, 10]. Committee-based approaches [9, 10] create an ensemble of classifiers using subsets of the training data and identify instances from an unlabeled pool which lead to maximum disagreement in prediction among the classifiers. Active Learning has been extensively studied for classification [7, 8, 11], where the goal is to identify unlabeled instances to be labeled according to membership in a class.

Active Learning for Collaborative Filtering is a relatively new concept; here the goal is to solicit ratings for minimal set of movies from a user to learn about the user's preference pattern. This is very essential because when new users join existing models, the system knows little about their preferences and the system would like to understand the user preference pattern with the least amount of training examples so as to not annoy the user with lot of questions.

Previous work on Active Learning applied to Collaborative Filtering [5, 12], have made one common implicit assumption, that users would be able to provide rating to any item that is requested by the system. This assumption, in reality, is not true, because, to rate a movie, a person has to first procure the movie, and watch it. This can be very

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR'08, July 20–24, 2008, Singapore.

Copyright 2008 ACM 978-1-60558-164-4/08/07 ...\$5.00.

time consuming, as opposed to, for example, in text classification where users can quickly label text documents just by skimming through the snippets. Moreover, users would not like a system which solicits ratings for movies the user may not even watch and such a dialog can be frustrating. Since Collaborative Filtering deals with user-personalization, we believe Active Learning should also be *personalized* to solicit ratings about items that a user would usually watch. In this paper, we provide one such *personalization* approach for active learning applied to Collaborative Filtering.

2. RELATED WORK

2.1 Collaborative Filtering

Aspect Model [2, 3] is a probabilistic latent semantic model in which users are considered to be a mixture of multiple interests or aspects. Thus each user $u \in U$ has a probabilistic membership in each of the aspects, $z \in Z$. Furthermore, users in the same interest groups have same movie-rating patterns, thus, users and items (e.g. movies $m \in M$) are independent from each other given the latent class variable, z . Each item m can be rated on a finite scale (e.g. $R = \{1, 2, \dots, 5\}$), with the rating $r \in R$. Thus the probability of each tuple in the dataset can be computed as follows:

$$P(r|u, m) = \sum_{z \in Z} p(r|m, z)P(z|u) \quad (1)$$

Equation 1 consists of two parts, $p(r|m, z)$ and $P(z|u)$. It can be observed that the first term $p(r|m, z)$ does not depend on the user and represents the group-specific model. The *global-model* consists of such group-specific models. The second term $P(z|u)$ is the user-personalization term. The *user-model* θ_u consists of such user-personalization terms i.e. $\theta_u = \{\theta_{u_z} : \theta_{u_z} = P(z|u), \forall z \in Z\}$.

The **Flexible Mixture Model** (FMM) [4], is a modified version of the Aspect Model in which there are two layers of latent aspects $\{z_u, z_m\}$, one grouping the users with similar interests (z_u) and one grouping the items with similar patrons (z_m). The probability of a tuple (r, u, m) , is factored similar to the Aspect Model:

$$P(r, u, m) = \sum_{z_u, z_m} P(z_u)P(z_m)P(u|z_u)P(m|z_m)P(r|z_u, z_m) \quad (2)$$

Both the Aspect Model and the Flexible Mixture Model are probabilistic clustering models and are fit to the data using Expectation Maximization (EM) [13, 2].

Users generally follow different distributions to rate the items. For example, some users provide very extreme ratings while some users provide moderate ratings. For learning models which generalize over such diverse users, the ratings are usually normalized so that ratings for each user have zero mean and unit variance [2].

2.2 Active Collaborative Filtering

When new users join a system, although the system already has a strong global-model, the user-model for new users is very weak, since the system has very few ratings available from such new users. Thus, the goal of active learning in CF is to obtain ratings for more items from the new

user. Instead of randomly selecting movies for rating from the user, active learning algorithms minimize the number of such queries required to learn a stronger user-model.

One of the popular techniques of active learning is to solicit a rating for an item which will minimize the expected entropy of the user model. In the context of Collaborative Filtering, this leads to the following equation for a user u , where θ_{u_z} denotes the user-group mixing probabilities $P(z|u)$, and $\theta_{u_z|m,r}$ denotes the model posterior after re-training the user-model based on a newly obtained rating r for movie m from the user i.e. $P(z|u, m, r)$:

$$m_u^* = \operatorname{argmin}_{m \in M} - \left\langle \sum_{z \in Z} \theta_{u_z|m,r} \log \theta_{u_z|m,r} \right\rangle_{P(r|u,m)} \quad (3)$$

Equation 3 denotes the expected entropy of the user-model after being trained over additional information of rating movie m with rating r . Since the exact rating r is not known for the unrated movies, the expected value of rating based on the current model $P(r|u, m)$ is used, as shown in the equation. Minimizing entropy leads to pure interest groups, in which each user has strict adherence to just one interest group. [12] demonstrate that minimizing entropy is counter-productive for active learning in the collaborative filtering domain, since in reality, users can be a mixture of multiple interests.

As a remedy to this problem, [12] propose a Bayesian Selection (BS) approach similar to the active learning approach towards parameter estimation in Bayesian networks [6]. This approach identifies item m , such that the updated model $\theta_{u|m,r}$ will be accelerated towards the true user model θ_u^{true} .

$$m_u^* = \operatorname{argmax}_{m \in M} \left\langle \sum_{z \in Z} \theta_{u_z}^{true} \log \frac{\theta_{u_z|m,r}}{\theta_{u_z}^{true}} \right\rangle_{P(r|u,m)} \quad (4)$$

Equation 4 is a negated KL-Divergence equation which is maximized when the estimated distribution is equal to the true distribution being modeled.

Since the *true* user model is unknown beforehand, it is estimated as the expectation over the posterior distribution of the user model. Similar to equation 3, the rating r is unknown for unrated movies and the expected value is used instead. [12] provides a computationally efficient way of approximating the posterior distribution.

3. PERSONALIZED ACTIVE LEARNING

The items selected by the Bayesian selection approach will accelerate the user-model to the *true* user-model, only if the user rates all the items which the system solicits ratings for. However, in reality, users cannot provide ratings for all items. For example, in reality, users do not watch all movies, do not buy/use all items, and hence cannot provide ratings for all items.

Thus, in addition to selecting the examples which will accelerate the model towards the true user model θ_u^{true} , the active learning algorithm should also try to select items which have a very high probability of getting a rating from a user. Thus we introduce a new term $P(m|u)$ into Equation 4, that is the probability of getting a rating, on the item m from the user u . Thus, the new objective function now becomes:

$$m_u^* = \underset{m \in M}{\operatorname{argmax}} \left(\left\langle \sum_{z \in Z} \theta_{u_z}^{\text{true}} \log \frac{\theta_{u_z|m,r}}{\theta_{u_z}^{\text{true}}} \right\rangle_{P(r|m,u)} \right) P(m|u) \quad (5)$$

The multiplication of the Bayesian Selection criterion from Equation 4 and the personalization term $P(m|u)$ acts as a soft-AND which is maximized when both the multiplicands are maximized. We demonstrate that this is a reasonable approach for Personalized active learning in our experiments.

We approximate the probability $P(m|u)$ as follows:

$$P(m|u) = \sum_{z \in Z} P(m|z)P(z|u) \quad (6)$$

$$P(m|z) = \frac{\sum_{u \in U} P(z|u)I(u, m)}{\sum_{u \in U} \sum_{m' \in M} P(z|u)I(u, m')} \quad (7)$$

$$I(u, m) = \begin{cases} 1, & \text{if user } u \text{ has rated item } m \\ 0, & \text{otherwise} \end{cases}$$

It should be noted that we use an indicator function $I(u, m)$ for computing $P(m|z)$ because, we are interested in computing the probability of getting any rating, instead of $P(r|m, z)$, i.e. the probability of getting a particular rating r for a movie m , in which case we would have used $P(r|u, m)$. It should also be noted that $P(m|u)$ is not computed as an expectation since this term is dependent on the user model only through $P(z|u)$ based on the current user-model.

Although our proposed method *extends* existing Bayesian active learning techniques for Aspect Model, it should be noted that our proposed method for personalization is general and intuitive and can be applied to any other Active Learning based Collaborative Filtering algorithm.

4. EXPERIMENTS

4.1 Experimental Setup

We use the MovieLens¹ and MovieRating² datasets for the empirical analysis of our approaches. The detailed characteristics of the two datasets are presented in Table 1.

The datasets were randomly split into training/test sets as shown in Table 2. The Aspect Model is first trained from training set. This gives the global model $p(r|m, z)$ as defined in Equation 1. Active learning involves soliciting ratings from new users. The test set forms this set of new users. Thus we start with very few initial ratings for each user, about 3, simulating a new user. The user-model θ_u is learnt from these ratings. For each test-user, remaining movies are split into two sets, the active-selection set and the evaluation set. Active learning algorithms select movies for rating from the set of active selection candidates. In previous approaches, the active-selection set consisted only of movies for which ratings are available in the dataset. This is unrealistic, since a system cannot know beforehand which movies will be rated by the user. This is also evident from the dataset characteristics reported in Table 1 in which

¹www.grouplens.org/system/files/ml-data.zip. This dataset is similar to the EachMovie dataset (prevalent in earlier literature) which is no longer available for usage. www.grouplens.org/node/76

²http://www.cs.usyd.edu.au/~irena/movie_data.zip

average number of movies rated by a user is less than 10% of the total number of movies. We use a more realistic setting, in which the active-selection set for each test-user consists of all the movies (rated and unrated) which are not included in 1) the initial-ratings and 2) the evaluation set for that user. This makes active-selection more challenging and the setting realistic. The evaluation set is a held-out set of movies for the test-user which is used for evaluating the performance of the Collaborative Filtering system.

We have set the number of user-classes to 5 and 10 respectively for the MovieRating and MovieLens datasets as reported in previous empirical studies and Active Collaborative Filtering literature [12].

4.2 Evaluation Metrics

We evaluate the system using two kinds of metrics:

CF performance: This includes the usual CF metrics, i.e. the Mean Absolute Error (MAE) and Mean Squared Error (MSE). MAE and MSE measure the deviation of the predicted ratings of the movies in the test set from the actual ratings available in the dataset. In the equations below, M_{ue} denotes the evaluation set for the user u and $n(M_{ue})$ denotes the number of movies in the set.

$$MAE_u = \frac{1}{n(M_{ue})} \sum_{m \in M_{ue}} |r_m^{\text{true}} - r_m^{\text{predicted}}| \quad (8)$$

$$MSE_u = \frac{1}{n(M_{ue})} \sum_{m \in M_{ue}} (r_m^{\text{true}} - r_m^{\text{predicted}})^2 \quad (9)$$

Since we use multiple test-users, the reported MAE (and MSE) is the average over individual MAE (and MSE) for each test-user as shown in Equation 8, where U_t is the set of test users and $n(U_t)$ denotes the number of test users.

$$MAE = \frac{1}{n(U_t)} \sum_{u \in U_t} MAE_u \quad (10)$$

Failures: The system solicits ratings for movies from the user and the user may not provide ratings for some of them. Such instances are known as *failures* to obtain rating from the user. This metric is very essential for comparison of the personalized approach with previous approaches, since it clearly identifies the number of queries from the active learning algorithms that will go unanswered, and thus not useful for the active learning system. High failure rate also means a system which is likely to annoy the user with questions for which the user does not have any answer. In our experimental setup, a failure occurs whenever the system solicits rating that does not occur in the dataset. The system thus cannot be re-trained and wastes an active-learning cycle and proceeds to the next iteration.

4.3 Active CF cycle

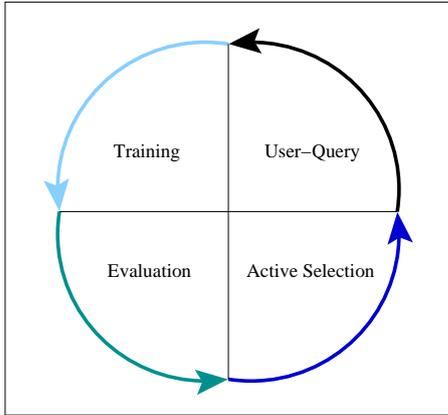
The Active learning cycle is shown in Figure 1. The global model is learnt over all the training-set users. The active learning cycle starts for each of the test users after the global model is learnt. The system learns a user-model for a test-user based on the available initial ratings, or the labeled-set and the system is then evaluated over the evaluation set. The system then selects a movie for rating from the active-selection set of the user. If the user provides rating for the movie, the system adds this movie-rating pair to

Table 1: Characteristics of MovieRating and MovieLens datasets

	MovieRating	MovieLens
Number of users	500	943
Number of movies	1000	1682
Average number of rated movies/user	87.7	106.05
Rating Scale	1-5	1-5

Table 2: Our Experimental Setup

	MovieRating	MovieLens
Number of Training-set users	200	343
Number of Test-set users	300	600
Number of Initial ratings per test user	3	3
Number of Preserved ratings (evaluation) per test user	20	20
Total number of Preserved ratings (evaluation) over all test users	6000	12000
Number of Active Selection Candidates per test user	977	1659
Number of user-classes	5	10

**Figure 1: The Active Learning process**

the labeled-set and retrain the system model. If the user does not provide rating for the movie, then it is noted as a failure. The system cannot be retrained in case of a failure. In either case the system is evaluated in terms of MAE and MSE on the evaluation set of the user. This completes one active-iteration. In the next step the system asks rating for another movie from the user and the cycle continues.

4.4 Aspect Model Implementation

For the Aspect Model, we model $P(r|m, z)$ with gaussian distributions $N(r; \mu_{m,z}, \sigma_{m,z})$. We use a multinomial model for the user-group mixing proportions $P(z|u)$, as proposed in [2]. Users generally do not follow the same distribution to provide ratings to items, and so it is common in the collaborative filtering techniques to normalize user-ratings to have zero mean and unit variance. It should be noted that the normalization parameters are learnt only over the ratings available to the system, not over the items in the active-selection candidates and evaluation set.

4.5 Comparative Baselines

The proposed Personalized Active Learning algorithm is compared against following approaches towards item selection:

- **Random Selection (RS)**: Items are selected randomly from the set of active-selection candidates for soliciting ratings from the user.
- **Bayesian Selection (BS)**: We implement the Bayesian approach outlined in [12]. This approach has already been compared to other existing approaches like *Model Entropy based Sample Selection*, which selects item based on expected reduction in model entropy, and *Prediction based Sample Selection* which selects items based on the uncertainty in prediction. According to [12] Bayesian Selection outperforms the other approaches.

We replicate the experiments in [12] with our implementation of the baseline in Figure 2. In our comparative evaluation, the baseline performs similar to [12] using the experimental setup reported in [12]. It should be noted however that [12] performs active-selection only on the subset of movies for which ratings are available in the dataset, thereby making the assumption that user can rate any movie. We call this setup *constrained* because the active-selection set is constrained to rated items. This is unrealistic, since the system will not know beforehand, which items will be rated by the user. Therefore, for comparison with *personalized Active Learning*, we will be using the entire set of movies, not just the ones for which ratings are available in the dataset. We call this the *unconstrained* setup.

For the sake of conciseness, we call our own approach the **Personalized Bayesian Selection (PBS)**.

4.6 Empirical Results and Discussion

The results for our experiments are presented in Figure 3, for the first 10 active iterations of the system.

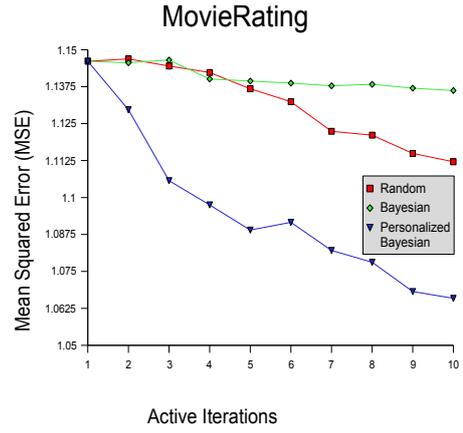
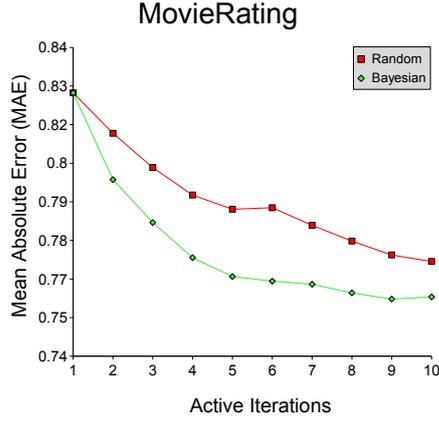
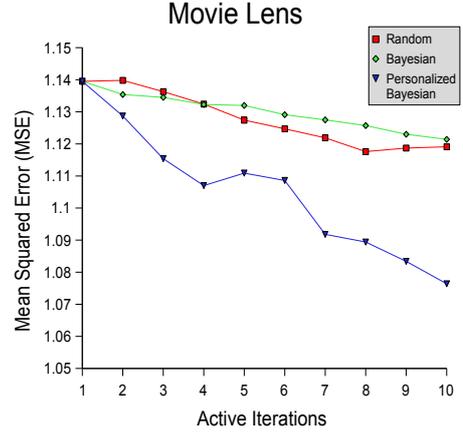
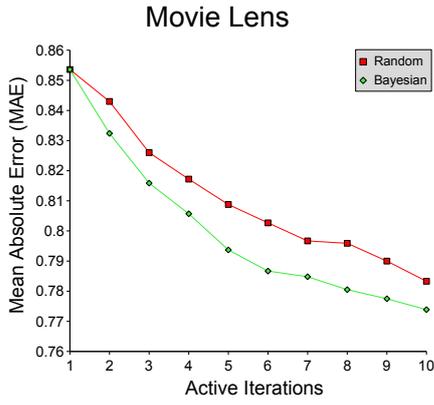


Figure 2: Constrained Baseline implementation: Active learning trends over 10 active-iterations

It can easily be observed that our approach, Personalized Bayesian Selection outperforms Bayesian and Random Selection on both datasets using both metrics, MAE and MSE. Moreover, since Bayesian Selection does not take the probability of getting a rating into account, it performs even worse as compared to plain Random Selection. This is expected because Bayesian selection tries to identify items that will provide the most information about user-model, but in the process often selects items which the user would not be able to provide a rating for. This is evident from the Table 3 where we report the mean number of failures, over all active-users for 10 active iterations.

Table 3: Mean number of Failures over 10 active iterations (averaged over 300 active-users for MovieRating and 600 active-users for MovieLens)

	MovieRating	MovieLens
RS	9.0728	9.3268
BS	9.8398	9.8366
PBS	5.995	5.1267

As shown in Table 1, average number of ratings per user is about 87.7 (out of 1000) and 106.05 (out of 1620) for the MovieRating and MovieLens datasets respectively, each user

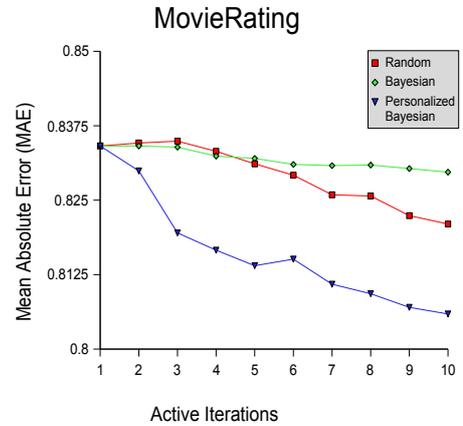
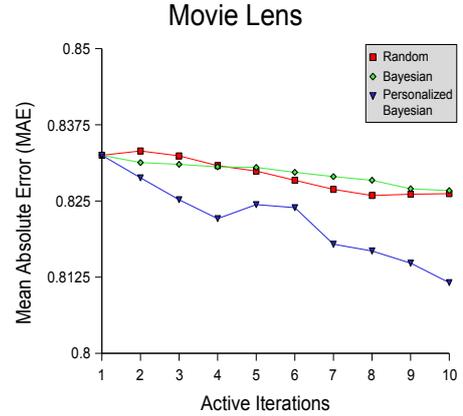


Figure 3: Unconstrained setup: Active Learning trends over 10 active-iterations

has on an average only 1 item rated out of 10. Thus, it is expected that random selection has about 9 failures out of 10 active-iterations. Bayesian Selection has more failures than Random Selection because most informative items may not be rated by the user. Personalized Bayesian selection substantially reduces the number of failures and selects at least 4 out of 10 items for which the active-user can provide a rating.

It should be noted that active CF systems will solicit ratings for new movies from the user till the system has substantial number of ratings from the user. A system with higher failure rate will tend to ask more questions to obtain comparable number of ratings from the user. Users typically do not favor systems which ask too many questions. Thus, a *personalized* active CF system is more likely to be favored by the user than an unpersonalized one.

4.7 Significance Tests

We conducted paired t-tests [14] for comparing the performance of the PBS, BS and RS approaches. Given two systems (say A and B) and a set of n items, the MAE (or MSE) values were computed for both systems, for each test item, denoted as a_i, b_i for $i = 1, 2, \dots, n$. We used the pairwise difference $d_i = a_i - b_i$ to examine the null hypothesis that expected difference $\bar{d} = \frac{\sum_i d_i}{n}$ is zero against the alternative hypothesis $\bar{d} > 0$, meaning system A is better than system B . The p-value is computed using the t-distribution with the degree of freedom $n - 1$ for

$$T \geq \frac{\bar{d}}{s.e.(\bar{d})} \quad (11)$$

Table 4: p-values for the significance tests comparing various active learning approaches

System A	System B	p-value	
		MovieRating	MovieLens
RS	BS	0.04	0.0019
PBS	BS	6.2e-05	2.1e-13
PBS	RS	0.0040	8.9e-08

Small p-value (≤ 0.01) means statistically significant evidence against the null hypothesis. Table 4 reports the p-values obtained in our tests using the MAE values of each system after 10 active learning iterations, i.e. after selecting 10 items and adding their ratings (if rated) to the current training set. It can be observed that our approach (PBS) significantly outperformed Bayesian Selection (BS) and Random Selection (RS). Random Selection performed better than Bayesian Selection with statistically moderate evidence, i.e., the p-values were less than 0.05 and above 0.01 for these comparisons.

4.8 Analyzing Personalization

The personalization term $P(m|u)$ in the active-selection Equation 7 consists of two terms, $P(z|u)$, the user-group mixing probabilities and $P(m|z)$, the probability of getting a rating for a movie m in group z .

In the Table 5, we present lists of movies in two exemplary interest-groups learnt for the MovieRating dataset. We show

Table 5: Exemplary movie clusters sorted on $P(m|z)$ for the MovieRating dataset

Cluster A
Dante's Peak
Contact
Scream
Air Force One
Murder at 1600
The Game
Conspiracy Theory
Titanic
I know what you did last summer
The Devil's own
Cluster B
Star Trek: First Contact
Return of Jedi
Scream
Courage under Fire
Contact
Mission Impossible
The Godfather
Star Trek: The wrath of Khan
The fifth element
Star Trek: Generations

top 10 movies in each group sorted in descending order of $P(m|z)$. It can be observed that cluster A consists of movies which thriller fans will view and cluster B consists of movies which usually science-fiction fans will view. If a user has watched some of the movies in a cluster, it is very likely that the user has also watched other movies in that interest group. Thus it is reasonable to ask the user to rate the other movies in that cluster to learn a stronger user-model. Unpersonalized methods do not follow such an approach and movies will be identified for rating from any of the interest groups, even from those in which the user has never watched any movies. This is the main reason for the higher failure rate of the unpersonalized approaches.

5. CONCLUSION AND FUTURE WORK

We have shown that personalization of active learning is essential for getting benefit from a Bayesian active learning approach to collaborative filtering. Personalization can be achieved in a relatively simple way: by selecting the most informative items for a user and the likelihood of the user to rate those items. Our experiments on benchmark datasets provide strong evidence for theoretical advantage of our approach over a baseline that is representative for well-established active learning methods without personalization in collaborative filtering: rate those items.

For future work, we plan to incorporate implicit feedback into the model. Thus, when a system selected item fails to get a rating, we can use failure as additional information to update the user-aspect probabilities $P(z|u)$. Google recently demonstrated techniques for making Aspect Model scalable to the number of news articles for collaborative-filtering based personalization of its news services [15]. A similar scalability study on Active Learning techniques for Collaborative Filtering is required, in addition to the scalability analysis of our personalized approach.

6. ACKNOWLEDGMENTS

This work is supported in parts by the National Science Foundation (NSF) under grant IIS-0434035 and IIS-0704689. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

7. REFERENCES

- [1] Breese, J. S., Heckerman, D. and Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering . Technical Report MSR-TR-98-12, 1998
- [2] Hoffman, T. Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis SIGIR 2003
- [3] Hofmann, T., and Puzicha, J. Latent class models for collaborative filtering. In Proceedings of International Joint Conference on Artificial Intelligence, 1999.
- [4] Si, L. and Jin, R. Flexible Mixture model for collaborative filtering. In Proceedings of the Twentieth International Conference on Machine Learning, 2003
- [5] Boutilier C., Zemel R., and Marlin B. Active Collaborative Filtering . SIGIR 2003
- [6] Tong, S. and Koller, D. Active Learning for Parameter Estimation in Bayesian Networks. In Advances in Neural Information Processing Systems, 2000.
- [7] Tong, S. and Koller, D. Support vector machine active learning with applications to text classification. In Journal of Machine Learning Research, Nov., 2001
- [8] McCallum, A. and Nigam, K. Employing pool-based active learning for text classification. In Proceedings of International Conference on Machine Learning, ICML-1998
- [9] Seung, H. S., Opper, M. and Sompolinsky, H. Query by Committee. In Computational Learning Theory, 1992
- [10] Freund, Y. Seung, Shamir, E. and Tishby. N. Selective sampling using the query by committee algorithm. Machine Learning, 1997
- [11] Godbole, S., Harpale, A., Sarawagi, S. and Chakrabarti, S. Document classification through interactive supervision of document and term labels. In ECML/PKDD, 2004
- [12] Jin, R. and Si, L. A Bayesian Approach toward Active Learning for Collaborative Filtering. In UAI, 2004.
- [13] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. In Journal of Royal Statistical Society, B39:1-38, 1977.
- [14] Yang, Y. and Liu, X. A re-examination of text categorization methods. In Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval, 1999
- [15] Das, A., Datar, M. and Garg, A. Google News Personalization: Scalable Online Collaborative Filtering. In WWW 2007.