Second-order Mining for Active Collaborative Filtering

Lingfeng Nia,b, Jianmin Wud, Yong Shia,b,c,*

aResearch Center on Fictitious Economy and Data Science, CAS, Beijing 100190, China
bGraduate University of Chinese Academy of Sciences, Beijing 100190, China
cCollege of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA
dYahoo! Research & Development (Beijing) Center

Abstract

Active learning for collaborative filtering tasks draws many attentions from the research community. It can capture the user’s interest with greatly reduced labeling burden for the online user. High quality recommendation can thus be made with good user experience. In this paper we address the efficiency challenge of current active learning methods for online and interactive applications by using the second-order mining techniques. According to the global latent semantic model learnt from the feedbacks of historical users to items, we propose an intuitive and efficient query strategy for the item selection for new active user. The time complexity in each query is reduced greatly to constant \(O(1)\). Experimental results on the public available data sets show the efficiency and effectiveness of our method.

Keywords:
Second-order Mining, Active Learning, Collaborative Filtering

1. Introduction

Active learning is a subfield of machine learning. It takes the hypothesis that if the learning algorithm is allowed to choose the data items from which it learns, it would perform better with less training examples. An active learner can ask queries in the form of unlabeled items to be labeled by an oracle, for example, a human annotator. It draws many attention from research communities especially for the learning problems that the label of example is difficult, time-consuming or expensive to be obtained, but the unlabeled data items are abundant. The key gradient in the active learning system is the strategy to select the query item to solicit the label from the oracle. The query is conducted by selecting the item with minimal loss defined by different strategies [1]. Different loss functions are defined to either maximize the expected information in the obtained model [2, 3] or to maximize the expected error reduction in prediction [4].

Collaborative Filtering (CF) is the technology to build recommender system based on the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users. Memory-based methods and model-based methods are the two major types of algorithms for CF. As the early generation of CF, such as GroupLens [5], memory-based methods use the users’ preferences data to calculate the similarity or weight between...
users or items. The recommendation or prediction are then made according to these calculated similarity values. The memory-based methods are easy to implement and highly effective in most of the time [5, 6]. However, due to the limitation of calculating the similarity based on users’ rating directly, the model based methods are sensitive to the data sparsity problem because the common items are few consequently. The model-based methods are therefore proposed to address this limitation. A machine learning model are estimated or learnt from the pure users’ preferences data. The recommendations or predictions for new user are then performed according to the learnt model. The well-known model-based CF algorithms include Bayesian belief networks CF methods [7, 8, 9], clustering based CF methods [10, 11], latent semantic CF methods [12, 6, 13], multiple-cause vector quantization (MCVQ) method for CF [14], and maximum-margin matrix factorization method [15, 16] etc..

Applying active learning to CF task draws many attention from the research community. The CF problem is considered in the interactive and online scenario. After obtaining a global model based on the corpus of users and their feedbacks to the whole item set, the online CF predicts the preference of new active users to items. For a new user with feedbacks on few items and the initial estimation of user’s interest distribution, the interactive and online CF procedure starts to solicit feedback of the new selected item based on the current user’s interest distribution. A new user’s interest distribution is then estimated with the current accumulated feedbacks on items. This procedure iterates until the change of user interest distribution less than given threshold or the limitation of feedbacks number being reached. The target of active learning in this procedure is to select the item that most improves the quality of recommendations or predictions made at each step and solicit feedback from user for this item. There are quite a lot of works conducted on this field [17, 18, 19]. The global model is utilized indirectly by the query strategies of these active learning methods. However, there are still several drawbacks in current existing methods [20]: 1) the query strategies are lack of intuitive explanation; 2) the intuitive explanation of the global model is not utilized directly in query strategies; 3) most of the query strategies are computationally expensive.

In this work, we propose a second-order mining method to address these drawbacks of the existing methods. Instead of using the information theory based criteria for sample selection, our method is primarily initiated according to the intuition in practice. By using the latent semantic model [21] as the underlying CF method, our approach selects the informative sample with the consideration of both the mean of ratings and the latent topic distribution.

The structure of this paper is organized as the follows. We review the related works on the active learning for CF in Section 2. The latent semantic model for CF and our second-order mining method for active learning are introduced in Section 3. We test our method by experiment on public data set in Section 4. In the last, the conclusion and future work are summarized in Section 5.

2. Related Works

In addition to the suggestion of active learning for CF in [22], the initial work [18] introduces the practical active learning method for CF. The CF problem is considered in the online and interactive scenario: given the current ratings provided by the active user, what query in the next rating would most improve the quality of recommendations made? The expected value of information is employed to select the item at each step to solicit the rating. The framework proposed in this paper works for most of the probabilistic model based CF methods, and multiple-cause vector quantization (MCVQ) [14] is used as an instance to illustrate the usage in this work.

In the work [17], the author proposed a Bayesian approach for the active learning in CF. In this work, the authors address the sparse and unstable problem of the estimated model during the active learning process. Most of the query strategies are framed according to the current estimation of the model to the obtain the item with minimal loss by some predefined loss function. However, at the initial stage, the estimated model could be very far from the optimal one due to the limited information available for the active user. The Bayesian method is employed to calculate the expected loss function over a posterior distribution of the model. Aspect model [12, 21] is used as the underlying model for CF. The posterior of the user interest model are approximated as a Direchlet distribution, which is obtained by expanding the posterior at the point obtained from maximal likelihood estimation. Experimental results show that the Bayesian method performs better than the entropy [2, 3] and prediction error reduction [4] based methods.

Efficiency is a major challenge in the previous active learning method for CF. At each round of query, inference of user interest should be performed for each rating of all the items in active learning pool. This is prohibitively time consuming for online and interactive applications. In [20], the author proposed an active learning method for
the CF with maximum-margin matrix factorization as the underlying model. The idea is initiated in the work [23] for classification problem with Support Vector Machines (SVM) as the classifier. The unlabeled items that are near to the separating plane and thus have less confidence are selected as the item to solicit label. These marginal items contribute more discriminative information to the SVM model when getting labeled and added to the training set.

All the aforementioned query strategies assume that the active user can provide feedback for any of the given item. This assumption, in reality, is not true. To provide feedback, the active user has to be familiar with the selected or even synthetic item [24]. This is impossible considering the large scale of active learning pool and the fact that potentially every item could be selected for soliciting the feedback. In [19], the authors proposed the personalized active learning process, the model of which considers the probability of the active user providing feedback to the given item. The Bayesian selection criteria in [17] is multiplied by the personalized term of the probability to obtain feedback from the active user.

However, we notice that to obtain the user preference over different type of items, it is not necessary to use very rare items. The popular and representative items for each category can provide us enough information about the active user. The challenges are to identify the underlying categories of items, which are not available in most of the case, and obtain the representative items for each category. In the following of this paper, we introduce how the underlying categories are discovered by the latent semantic model, and how we obtain the representative items for each category.

3. Second-order mining for active CF

3.1. Latent semantic model

Latent semantic model [12, 21] is one of the model-based method for collaborative filtering. Let $\mathcal{U} = \{u_1, \cdots, u_m\}$ denote a set of users and $\mathcal{Y} = \{y_1, \cdots, y_n\}$ denote a set of items. The co-occurred triplet $(u, y, r)$ records that the preference rating $r$ is given to item $y$ by user $u$. The rating score $r \in \mathbb{R}$ shows the user’s preference judgement on the item.

Latent semantic model takes the assumption that there are latent semantics or topics hidden behind these user ratings. There are a number of ways to extend the dependency structure of the dyadic co-occurrence [25] to include an additional rating variable [12]. The extension that has been empirically proven to be most useful is to introduce direct dependencies of the rating variable on the item in question, but to mediate the dependency on the user through a latent variable [12]. The dependency of user on the latent variable is supposed to capture user communities or interest groups. The conditional probability $p(r\mid u, y)$ of the observed triplet can then be represented by the following mixture model:

$$p(r\mid u, y) = \sum_z p(r\mid y, z)p(z\mid u).$$

The rating $r$ here is not categorical but numerical variable. It is thus not appropriate to use multinomial distribution to parameterize the model. Gaussian distribution is used instead for the parameter form of $p(r\mid y, z)$, and multinormal distribution is used for $p(z\mid u)$. The resulting distribution $p(r\mid u, y)$ is thereby a user-specific mixture of Gaussian distribution. Formally, let us use $\mu_y$ and $\sigma_y$ to represent the mean and standard deviation for the Gaussian distribution, the parameter form of $p(r\mid u, y)$ is thereby

$$p(r\mid u, y) = \sum_z p(r; \mu_y, \sigma_y)p(z\mid u),$$

where $p(r; \mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2\sigma^2}(r - \mu)^2\right\}$. It is known that different users may associate subjectively different meanings with ratings. For example, rating score 3 for a tough user may refers to the best item, but the same score may means differently for a tender user. Therefore, all the ratings distribution for user $u$ are normalized to the Gaussian distribution with mean zero and variance one by

$$(u, r, y) \leftrightarrow (u, r', y) : r' = \frac{r - \mu_u}{\sigma_u}. $$

With the mixture distribution, the expected rate for user $u$ on item $y$ can be calculated by

$$E[r\mid u, y] = \int_{\mathbb{R}} rp(r\mid u, y)dr = \sum_z \mu_y p(z\mid u).$$
For the given users and their ratings over the item set, model parameters $\theta = (\mu_{y,z}, \sigma_{y,z}, P(z|u))$ are obtained by the maximal likelihood estimation over the observed rating set $D = \{(r, u, y)\}$. The log-likelihood is defined as

$$L(D, \theta) = \sum_{(r, u, y)} \log p(r|u, y; \theta).$$  

(3)

Expectation Maximization (EM) algorithm is employed to obtain the local maximal likelihood [21]. Each step of the EM training process consists of the E-step and M-step, which can be summarized as follow for the latent semantic model:

- **E-step:** estimate the posterior distribution by

$$P(z|r, u, y; \hat{\theta}) = \frac{\hat{p}(r|y, z)\hat{P}(z|u)}{\sum_{z'} \hat{p}(r|y, z')\hat{P}(z'|u)}.$$  

(4)

here we use $\hat{\theta}$ to represent the current estimation about the parameters $\theta$. The hat on the right hand side of equation (4) indicates the quantities parameterized by $\hat{\theta}$.

- **M-step:** update the user interests parameter $P(z|u)$ in $\theta$ by the equation

$$P(z|u) = \frac{\sum_{<r, u, y> | y = z} P(z|r, u, y)}{\sum_{z'} \sum_{<r, u, y> | y = z'} P(z'|r, u, y)},$$  

(5)

and parameters of Gaussian distribution ($\mu_{y,z}, \sigma_{y,z}$) by equations

$$\mu_{y,z} = \frac{\sum_{<r, u, y> | y = z} r P(z|r, u, y)}{\sum_{<r, u, y> | y = z} P(z|r, u, y)}$$  

(6)

$$\sigma_{y,z}^2 = \frac{\sum_{<r, u, y> | y = z} (r - \mu_{y,z})^2 P(z|r, u, y)}{\sum_{<r, u, y> | y = z} P(z|r, u, y)}.$$  

(7)

The EM algorithm iterates by E-step and M-step alternatively until a local maximal likelihood value is reached. The obtained optimal parameters $\mu_{y,z}$ and $\sigma_{y,z}$ for the items are then served as the global model for inferring the interests of new active users.

### 3.2. CF with latent semantic model

After obtaining the global model, the active learning process for collaborative filtering is to first select the item to solicit rating from the user according to his/her interests distribution. Based on the new user rating over the selected item and the current accumulated ratings for the same user, we infer the new interest distribution for this user. The quality of new interest distribution is then evaluated with certain measurement. This process is repeated until some required criteria is met. We show the overall process in Algorithm 1. Inference of user interest distribution in the step 3 in Algorithm 1 is done by the same EM procedure but keeping the global parameters $\mu_{y,z}$ and $\sigma_{y,z}$ fixed during the iteration process.

As we noticed in Section 2 that for the entropy, prediction or Bayesian methods [2, 4, 17] for active learning, this inference process should be done for each ratings of all the items in the active learning pool. For each active user, the time complexity for each query is $O(nRK)$. Although it can be accelerated by the posterior distribution of user’s interests [17], the query process is still not suitable for the online and interactive applications [20]. We introduce our method to address this challenge by the second-order mining method in the rest of this paper.

### 3.3. Second-order mining query strategy with latent semantic model

For the active user $u$, we aim at obtaining the true interests distribution $P^*(z|u)$ with as less as possible queries for soliciting ratings for selected items. Previous works for query strategies in active learning are designed to minimize the loss function defined over either the obtained model itself [3, 26, 27] or the prediction error of the new model [2, 4]. On the other hand, many of the global model parameters obtained from the training set are framed with
Algorithm 1 Active Learning Process for CF

Require: Global model parameters $\mu_{y,z}$ and $\sigma_{y,z}$; The active learning pool $Y$; Active user set $U$ for recommendation;

for Each user $u \in U$ do
    Initialize the rating seeds $S$;
    repeat
        1) Select the item $y$ from the pool $Y$ based on the query strategy and current estimation of the user interest $P(z|u)$;
        2) Solicit the rating $r$ for item $y$ from user $u$; add $(u, y, r)$ to rating set : $S := S \cup \{(u, y, r)\}$;
        3) Infer the user interest distribution according to $S$;
        4) Evaluate the quality of user interest distribution $P(z|u)$;
    until Required precision achieved
end for

intuitive explanations. For the Gaussian latent semantic model, in addition to the user interests distribution $P(z|u)$, the mean $\mu_{y,z}$ of Gaussian distribution represents the average ratings of item $y$ over topic $z$ given by users, whereas $\sigma_{y,z}$ reflects the variance of these ratings. Generally speaking, the larger the mean $\mu_{y,z}$ is, the more representative the item $y$ is in topic $z$.

Our new query strategy is initiated by the following observation. The user interests or communities can be determined efficiently by soliciting ratings for the most representative items for each topic. Let us take the movie rating as an example, to check if the user $u$ is interested in romantic movie, we can solicit the ratings for representative movies like Titanic, Notting Hill, etc., in this topic. From the obtained ratings, we then infer the preference of user $u$ to the romantic type of movies. The new method to select the item for rating also alleviates the problem of soliciting ratings for items that user may cannot provide rates [19]. This is because representative movies in each topic are usually the well known ones. The active user most likely watched or at least heard about the movies compared with the non-representative and rare movies.

To be concrete, for each topic $z$, we sort items related to this topic in the ascent order of $\mu_{y,z}$. Then for the active user $u$, at each round of query, we select the item from the sorted list for each topic alternatively to solicit the rating. The process is illustrated in Figure 1. As we can see from this process, the complexity of the our query strategy mainly consists of the sorting for each topic, which is much less than previous entropy or Bayesian methods[3, 17]. The complete procedure of the second-order method is summarized in Algorithm 2. We test the efficient and effectiveness of the new method in the following experiment section.

Figure 1: Illustration of the items order to solicit ratings

3.4. Complexity analysis of second-order mining query strategy

In addition to the intuitive explanation of the query strategy with second-order mining method, the time complexity is still reduced significantly in the new query strategy. The sorting of items according to $\mu_{y,z}$ in each latent topic $z$ can be done before the active learning procedure for each new user. The time complexity for the sorting is $O(Kn \log n)$. The time complexity of the query strategy for each active user is $O(1)$, which is independent of the cardinality of ratings as well as the size of active learning pool, and thus suitable for online and interactive applications.
Algorithm 2 Second-order Mining for Active CF

Require: Global model parameters $\mu_{y,z}$ and $\sigma_{y,z}$; The active learning pool $\mathcal{Y}$; Active user set $\mathcal{U}$ for recommendation;
for $z = 1 : K$ do
  Sort items for topic $z$ in the descending order of $\mu_{y,z}$;
end for
for Each user $u \in \mathcal{U}$ do
  Initialize the rating seeds $\mathcal{S}$.
  repeat
    1) Select the item $y$ from the pool $\mathcal{Y}$ in the consecutive order of topic $z$ and sorted order of items list;
    2) Solicit the rating $r$ for item $y$ from user $u$; add $(u, y, r)$ to rating set $\mathcal{S} := \mathcal{S} \cup \{(u, y, r)\}$;
    3) Infer the user interest distribution according to $\mathcal{S}$;
    4) Evaluate the quality of user interest distribution $P(z|u)$;
  until Required precision achieved
end for

Table 1: Data statistics

|                  | MovieRatings | MovieLens |
|------------------|--------------|-----------|
| # Users          | 500          | 943       |
| # Movies         | 1,000        | 1,682     |
| Average # movies rated per user | 87.7         | 106.1     |
| Ratings set      | \{1,2,3,4,5\} | \{1,2,3,4,5\} |

4. Experiments

4.1. Data sets

Two public available data sets about movie ratings are used to test the effectiveness and efficiency of our method: MovieRatings$^1$ and MovieLens$^2$. The MovieRatings data include the movie ratings collected through user interactions with the site www.movielens.org. This includes ratings of user to movies on the scale of 1(worst) to 5(best). The MovieLens data set was collected by the GroupLens research project at University of Minnesota. The selected data consists of 100,000 ratings from 943 users to 1,882 movies with the same scale of ratings as data set MovieRatings. The detailed statistics of each data set are shown in Table 1. The same data sets are also used in [17] and [19] for the evaluation.

4.2. Settings and measurements

We use the first 200 and 343 users in data set MovieRatings and MovieLens for the global model training. The number of latent topics is set to 5 and 10 for MovieRatings and MovieLens, respectively. All the remaining users are used to test the effectiveness of our method. The bayesian method in [17] is employed as the baseline. For each testing user, we randomly select 3 movies from the rated movies set as seeds to obtain the initial estimation of the user’s interests. 20 movies are used as the evaluation set. The remaining ones are served as the active selection pool. For each user, we select up to 5 movies one by one and solicit the rate. The same settings are used in [17] and [19]. The evaluation is done at each step with the Mean Absolute Error(MAE) as the measurement:

$$MAE = \frac{1}{N_{test}} \sum_u \sum_n |r_n - E[r|u, y_n]|,$$

where $N_{test}$ is the number of rates involved in evaluation.

---

$^1$http://sydney.edu.au/~irena/movie_data.zip
$^2$http://www.grouplens.org/system/files/ml-data_0.zip
4.3. Effectiveness of the heuristic query strategy

We use the Bayesian method proposed in [17] as the baseline method. In the experiment, for each data set, we run the two methods three times. The mean and standard variance are shown in Figure 2 for the two data sets, respectively. From the figures, we can see that our heuristic based query strategy performs similar and even better than the Bayesian method. However, in terms of the time required for the query strategies, our method shows significant improvement. For the MovieLens data set, the Bayesian method takes 1,424 seconds, but our heuristic method takes 8 seconds on the same data set. For the MovieRatings data set, the Bayesian method takes 494 seconds, and our method takes 2 seconds.

![Figure 2: MAE of MovieRatings(Left) and MovieLens data sets](image)

4.4. Convergence of global model training

We show the convergence of the training for global model with the data set MovieRatings and MovieLens in Figure 3. As we can see, the training of the global model on the two data set converges with less than 50 iterations. Because the mixture of Gaussian distribution is used for parameterizing \( p(r|u, y) \), the log-likelihood formula (3) may result in positive values, which is a bit different from the classical log-likelihood definition.

![Figure 3: Log-likelihood of MovieRatings(Left) and MovieLens(Right) data sets to # iterations](image)
5. Conclusion and future works

We proposed a second-order mining based method for the task of active learning with CF in this paper. Our method exploits the intuitive explanation of the learnt global model. In addition to the intuitive explanation, our query strategy reduces the complexity of item selection procedure significantly, which enables it be suitable to the online and interactive applications. The same and even better quality of recommendation result is achieved compared with the Bayesian method.

One of the future works is to consider the extension of the idea to the case where the category information is directly observable from the collected data, which can improve the topic quality from latent semantic model. Combination of various type of statistics is also an interesting direction to explore. For example, equipping the mean with variance to address the problem of the selection of item with large mean but also large variance. The representative items for each topic would be more clean in this case.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (Grant No. 11026187, 70921061), and the BHP Billiton Cooperation of Australia.

References

[1] B. Settles, Active learning literature survey, Computer Sciences Technical Report 1648, University of Wisconsin-Madison (2009).
[2] S. Tong, D. Koller, Active Learning for Parameter Estimation in Bayesian Networks, in: NIPS, 2000, pp. 647–653.
[3] H. S. Seung, M. Opper, H. Sompolinsky, Query by committee, in: Proceedings of the fifth annual workshop on Computational learning theory, COLT ’92, ACM, New York, NY, USA, 1992, pp. 287–294.
[4] N. Roy, A. McCallum, Toward optimal active learning through sampling estimation of error reduction, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML ’01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001, pp. 441–448.
[5] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: an open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM conference on Computer supported cooperative work, CSCW ’94, ACM, New York, NY, USA, 1994, pp. 175–186.
[6] T. Hofmann, Latent semantic models for collaborative filtering, ACM Trans. Inf. Syst. 22 (2004) 89–115.
[7] J. S. Breese, D. Heckerman, C. Kadie, Empirical Analysis of Predictive Algorithms for Collaborative Filtering, in: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98), 1998, pp. 43–52.
[8] K. Miyahara, M. J. Pazzani, Collaborative filtering with the simple bayesian classifier, in: Proceedings of the 6th Pacific Rim international conference on Artificial intelligence, PRICAI’00, Springer-Verlag, Berlin, Heidelberg, 2000, pp. 679–689.
[9] X. Su, T. M. Khoshgoftaar, Collaborative filtering for multi-class data using belief nets algorithms, in: Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence, ICTAI ’06, IEEE Computer Society, Washington, DC, USA, 2006, pp. 497–504.
[10] L. Ungar, D. Foster, Clustering Methods For Collaborative Filtering, in: Proceedings of the Workshop on Recommendation Systems, AAAI Press, Menlo Park California, 1998.
[11] S. H. S. Chee, J. Han, K. Wang, Rectree: An efficient collaborative filtering method, in: Proceedings of the Third International Conference on Data Warehousing and Knowledge Discovery, DaWaK ’01, Springer-Verlag, London, UK, 2001, pp. 141–151.
[12] T. Hofmann, J. Puzicha, Latent class models for collaborative filtering, in: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI ’99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999, pp. 688–693.
[13] H. P. Ibrahim Yakut, Privacy-preserving svd-based collaborative filtering on partitioned data, International Journal of Information Technology & Decision Making 9 (2010) 473–502.
[14] D. A. Ross, R. S. Zemel, Multiple cause vector quantization, in: S. T. S. Becker, K. Obermayer (Eds.), Advances in Neural Information Processing Systems 15, MIT Press.
[15] N. Srebro, J. D. M. Rennie, T. S. Jaakkola, Maximum-Margin Matrix Factorization, in: Advances in Neural Information Processing Systems 17, Vol. 17, 2005, pp. 1329–1336.
[16] J. D. M. Rennie, N. Srebro, Fast maximum margin matrix factorization for collaborative prediction, in: Proceedings of the 22nd international conference on Machine learning, ICML ’05, ACM, New York, NY, USA, 2005, pp. 713–719.
[17] R. Jin, L. Si, A bayesian approach toward active learning for collaborative filtering, in: Proceedings of the 20th conference on Uncertainty in artificial intelligence, UAI ’04, AUAI Press, Arlington, Virginia, United States, 2004, pp. 278–285.
[18] C. Bourlier, R. S. Zemel, B. Marlin, Active collaborative filtering, in: In Proceedings of the Nineteenth Annual Conference on Uncertainty in Artificial Intelligence, 2003, pp. 98–106.
[19] A. S. Harpale, Y. Yang, Personalized active learning for collaborative filtering, in: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR ’08, ACM, New York, NY, USA, 2008, pp. 91–98.
[20] G. T. Irina Rish, Active collaborative prediction with maximum margin matrix factorization, in: The Tenth International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida, 2008.
[21] T. Hofmann, Collaborative filtering via gaussian probabilistic latent semantic analysis, in: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, SIGIR ’03, ACM, New York, NY, USA, 2003, pp. 259–266.
[22] 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 16th Conference on Uncertainty in Artificial Intelligence, UAI ’00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000, pp. 473–480.
[23] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, Journal of Machine Learning Research 2 (2002) 45–66.
[24] D. Angluin, Queries revisited, Theory Computation Science 313 (2004) 175–194.
[25] T. Hofmann, Probabilistic latent semantic analysis, in: Proceedings of Uncertainty in Artificial Intelligence, UAI, Stockholm, 1999.
[26] Y. Freund, H. S. Seung, E. Shamir, N. Tishby, Selective sampling using the query by committee algorithm, Machine. Learning 28 (1997) 133–168.
[27] C. Campbell, N. Cristianini, A. J. Smola, Query learning with large margin classifiers, in: Proceedings of the Seventeenth International Conference on Machine Learning, ICML ’00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000, pp. 111–118.