Sensitivity of Attributes on the Performance of Attribute-Aware Collaborative Filtering

Karen H. L. Tso and Lars Schmidt-Thieme

Computer-based New Media Group (CGNM),
Institute of Computer Science,
University of Freiburg, 79110 Freiburg, Germany
{tso,lst}@informatik.uni-freiburg.de

Abstract. Collaborative Filtering (CF), the most commonly-used technique for recommender systems, does not make use of object attributes. Several hybrid recommender systems have been proposed, that aim at improving the recommendation quality by incorporating attributes in a CF model.

In this paper, we conduct an empirical study of the sensitivity of attributes for several existing hybrid techniques using a movie dataset with an augmented movie attribute set. In addition, we propose two attribute selection measures to select informative attributes for attribute-aware CF filtering algorithms.

1 Introduction

For recommender systems, nearest-neighbor methods, called CF (Goldberg et al. (1992)), is the prevalent method in practice. On the other hand, methods that regard only attributes and disregard the rating information of other users, are commonly called the Content-Based Filtering (CBF). They have shown to perform very poorly. Yet, attributes usually contain valuable information that could improve the performance of recommender systems; hence it makes it desirable to include attribute information in CF models – so called hybrid collaborative/content-based filtering methods.

Although there are several hybrid methods that consider attribute information in CF for predicting ratings — how much a given user will like a particular item; to our best knowledge there is no prior approach for predicting items — which N items a user will be interested in. Please note that predicting good items, i.e. items that have been rated with 4 or 5 on a scale from 1 to 5, by its nature is a rating prediction problem (on a more coarse scale bad/good).

In addition, the behavior of hybrid algorithms is to be investigated as the number of informative attributes increases. Thus, quantitative measures for attribute selection are needed to eliminate irrelevant ones.

In this paper, we will make the following contributions: (i) propose two methods for attribute selection and (ii) evaluate the impact of attributes on existing hybrid algorithms that predict items and CF methods that do not consider attributes.

2 Related Work

Before we discuss the related works, we introduce notations being used in this paper. Let

- U be a set of **users**,
- I be a set of **items**,
- B be a set of (binary) **item attributes**,
- $D_{i,b} \in \{0, 1\}$ specify whether item $i \in I$ has attribute $b \in B$,
- $O_{u,i} \in \{0, 1\}$ specify whether item $i \in I$ occurred with user $u \in U$

There are many proposals on how to integrate attributes in collaborative filtering for ratings. They can be roughly categorized into four groups:
(i) Methods that add a pseudo-item i_b for each item attribute $b \in B$ that for each user $u \in U$ gets a pseudo-rating

$$R_{u,i_b} := f(\{(i, R_{u,i}) \in I \,|\, O_{u,i} = 1 \text{ and } D_{i,b} = 1\})$$

where f is some function on the user's ratings of items having attribute b. Ziegler et al. (2004) presented a more complex function that considers a taxonomic relation between original items.
(ii) Methods that add a pseudo-user u_b (often called agent) for each item attribute $b \in B$ with a pseudo-rating for each item $i \in I$

$$R_{u_b,i} := D_{i,b}$$

e.g., Good et al. (1999). These methods perform standard user- or item-based CF on top of the rating matrix enriched by pseudo-items or -users.
(iii)a) Methods that combine linearly the predictions of a pure CBF model and a pure CF model (Claypool et al. (1999), Pazzani (1999), Good et al. (1999), Li and Kim (2003))

$$\hat{R}^{\text{combined}} := \lambda \hat{R}^{\text{cbf}} + (1 - \lambda)\hat{R}^{\text{cf}}$$

where the weight coefficient $\lambda \in [0, 1]$ is learned either by regression, simple iterative update schemes or grid search. Some other existing methods also use a user-specific λ.
b)Apply the nearest neighbor models to both models and combining the attribute-depended with the rating-depended similarity and use CF with the combined similarity (Delgado et al. (1998)).
(iv) Methods that apply a CBF and a CF model sequentially, i.e. predict ratings by means of CBF and then re-estimate them from the completed rating matrix by means of CF (Melville et al. (2002)).

There are also further proposals on how to integrate attributes when the problem is viewed as a classification problem (Basilico and Hofmann (2004), Basu et al. (1998)). As we lose the simplicity of CF, we do not consider those more complex methods here.

Many methods appear to mix simple ideas with more complex components as clustering, rule-based learners etc., often without investigating whether the additional effort pays off in the quality at the end. Therefore, we have selected three basic methods that try to keep the simplicity of CF, but still should improve prediction results: a sequential CBF-CF method (iv), a linear-combination CBF-CF method (iii a) and a combination of similarities methods (iii b). The first approach is an adapted form of Content-Boosted CF by Melville et al. (2002) which was originally designed for predicting the ratings. The last two methods achieve the best results on our reference data set.

3 Common and Hybrid Attribute-Aware CF Methods

3.1 Common CF Methods

In user-based CF (Sarwar et al. (2000)), recommendations are generated by considering solely the ratings of users on items, by computing the pairwise similarities between users, e.g., by means of vector similarity

$$\text{usim}^{\text{ratings}}(u,v) := \frac{\langle R_{u,.}, R_{v,.}\rangle}{||R_{u,.}||_2 ||R_{v,.}||_2}$$

where $u, v \in U$ are two users and $R_{u,.}$ and $R_{v,.}$ the vectors of their ratings. For each user, the k most-similar users are selected (neighborhood – N_u) and for predicting items for a target user u, items are ranked by decreasing frequency of occurrence in the ratings of his/her neighbors

$$p^{\text{cf}}(O_{u,i} = 1) := \frac{|\{v \in N(u) \,|\, O_{v,i} = 1\}|}{|N_u|}$$

A dualistic form of user-based CF is item-based CF (Deshpande and Karypis (2004)), where similarities are computed between each pair of items $i, j \in I$.

$$\text{isim}^{\text{ratings}}(i,j) := \frac{\langle R'_{.,i}, R'_{.,j}\rangle}{||R'_{.,i}||_2 ||R'_{.,j}||_2}$$

In content-based filtering, a naive Bayesian classifier is trained for the binary target variable $O_{u,.}$ depending on the binary predictors $D_{.,b}$ for all $b \in B$:

$$\hat{p}^{\text{cb}}(O_{u,.} = 1 \,|\, D_{.,b}, b \in B) := P(O_{u,.}) \cdot \prod_{b \in B} P(D_{.,b} \,|\, O_{u,.}) \qquad (1)$$

3.2 Hybrid Attribute-aware CF Methods

The three existing hybrid methods in Tso and Schmidt-Thieme (2005) incorperate attributes into user-based and item-based CF.

Sequential CBF and CF is the adapted version of an existing hybrid approach, Content-Boosted CF, originally proposed by Melville et al. (2002) for predicting ratings. This method has been conformed to the predicting items problem here. It first uses CBF to predict ratings for unrated items and then filters out ratings with lower scores (i.e. keeping ratings above 4 on a 5-point scale) and applies CF to recommend topN items.

Joint Weighting of CF and CBF, first applies CBF on attribute-dependent data to infer the fondness of users for attributes. In parallel, user-based CF is used to predict topN items with ratings-dependent data. Both predictions are joint by computing their geometric mean. This mean combination is then used for performing the prediction:

$$\hat{p}(O_{u,i}=1) := \hat{p}^{\mathrm{cb}}(O_{u,i}=1)^{\lambda} \cdot p^{\mathrm{cf}}(O_{u,i}=1)^{1-\lambda} \quad \text{with } \lambda \in [0,1]. \tag{2}$$

Attribute-Aware Item-Based CF extends item-based CF (Deshpande and Karypis (2004)). It exploits the content/attribute information by computing the similarities between items using attributes thereupon combining it with the similarities between items using ratings-dependent data.

$$\begin{aligned} \mathrm{isim}^{\mathrm{attributes}}(i,j) &:= \frac{\langle D_{i,.}, D_{j,.}\rangle}{||D_{i,.}||_2 ||D_{j,.}||_2} \\ \mathrm{isim}^{\mathrm{combined}} &:= (1-\lambda)\ \mathrm{isim}^{\mathrm{ratings}} + \lambda\ \mathrm{isim}^{\mathrm{attributes}} \quad \text{with } \lambda \in [0,1] \end{aligned}$$

The last two methods use λ as weighting factor to vary the significance of CF or CBF.

3.3 Attribute selection

To our best knowledge, no similar analysis has been documented in literature affiliated with the sensitivity of attributes in RSs. As the number of attributes increases, quantitative measure for attribute selection are needed to filter the irrelevant ones. Thus, we define two quantitative measures (i) the total number of attribute occurrences (attribute frequency) and (ii) the χ^2 measure between item occurrences and attributes. The attribute frequency is simply the total number of occurrences of each item having a particular attribute. The more frequent an attribute is, the better it is judged.

This approach appears to be simple and clearly favors ubiquitous but attributes could eventually became non-informative. Thus, we also consider the χ^2 measure between item occurrences and attributes. It bases on the 2×2 table of all possible ratings $U \times I$ according to actual occurrence in the data $(O_{u,i})$ and having the attribute in question $(D_{i,b})$

	$D_{.,b}=1$	$D_{.,b}=0$
$O_{.,.}=1$	$v_{1,1}$	$v_{1,2}$
$O_{.,.}=0$	$v_{2,1}$	$v_{2,2}$

where

- $v_{1,1} := |\{(u,i) \in U \times I \,|\, O_{u,i} = 1 \text{ and } D_{i,b} = 1\}|$,
- $v_{1,2} := |\{(u,i) \in U \times I \,|\, O_{u,i} = 1 \text{ and } D_{i,b} = 0\}|$,
- $v_{2,1} := |U| \cdot |\{i \in I | D_{i,b} = 1\}| - v_{1,1}$,
- $v_{2,2} := |U| \cdot |\{i \in I | D_{i,b} = 0\}| - v_{1,2}$

computed by

$$\chi^2 = \sum_{i=0,1,j=0,1} \frac{(v_{i,j} - \hat{v}_{i,j})^2}{\hat{v}_{i,j}} \tag{3}$$

where $\hat{v}_{i,j} := v_{i,.} \cdot v_{.,j} / v_{.,.}$ represents the expected frequencies.

The stronger the dependency between item occurrence and an attribute, i.e. the higher the χ^2 value is, the better the attribute is judged.

4 Evaluation and Experimental Results

We have evaluated the three attribute-aware CF algorithms and have compared their performances with their corresponding non-hybrid base models, which do not integrate attributes.

Data set We evaluated the algorithms with the MovieLens datasets (*ml*; MovieLens (2003)), which contains approximately 1 million movie ratings of 6,040 users on 3,592 movies. The ratings are expressed on a 5-point rating scale. We looked at two different sets of movie attributes: (i) 18 genres that comes with the data set and (ii) Amazon taxonomy of 1074 different genres/classes provided by Ziegler et al. (2004). We will reference these two attributes sets as "18 genres" and "Amazon genres", respectively.

We took ten random subsets of the *ml* dataset with 1000 users and 1500 items each. Each dataset is split into 80% training set and 20% test set at random. The quality of the models are measured by comparing their top 10 recommendations computed from the training data against the actual items in the test set. We report the averages and standard deviations of the F1 values of the ten trials.

Metrics Our paper focuses on the item prediction problem, which is to predict a fixed number of top recommendations and not the ratings. Suitable evaluation metrics for item prediction problem are Precision, Recall and F1. Similar to Sarwar et al. (2000), our evaluations consider any item in the recommendation set that matches any item in the test set as a "hit".

Parameters We select optimal neighborhood sizes, by means of a grid search. Neighborhood size for user-based and joint weighting CF–CBF is 90, and 100 for item-based CF and attribute-aware item-based CF. Furthermore, λ parameters are set to 0.15 and 0.05 for joint weighting CF–CBF and attribute-aware item-based CF respectively. They are chosen from previous experiments (Tso and Schmidt-Thieme (2005)), which found to give reasonable results for the augmented attributes as well.

Experimental Results The results of our previous experiments (Tso and Schmidt-Thieme (2005)) on the 18 genres attribute set is summarized in Fig. 1. The attribute-aware methods enhance their respective base-models significantly, especially the joint weighting CF-CBF. Although Melville et al. (2002) reported that CBCF performed better than user-Based and CBF for ratings, it fails to provide quality topN recommendations for items in our experiments.

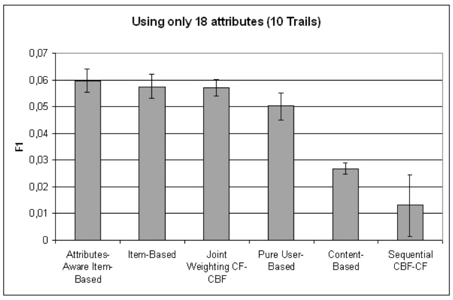

Fig. 1. F1 with 18 attributes

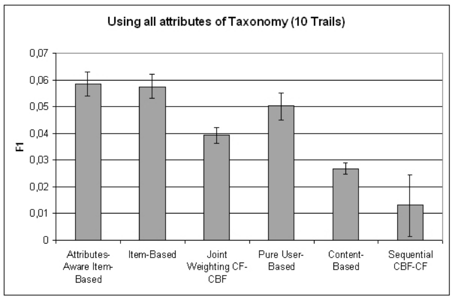

Fig. 2. F1 with all taxo attributes

We anticipate that the prediction quality could be improved by including more attributes. i.e. using the Amazon attribute set instead of the 18 genres. The results of the average of ten random trials using all attributes from the Amazon taxonomy are presented in Fig. 2.

Although attribute-aware item-based CF using all Amazon attributes still achieves the highest F1 value, the difference w.r.t. its base method is insignificant. It also can be observed that all attribute-aware methods perform worse for the 1074 Amazon attributes than for just the 18 genres. This indicates that the quality of attributes plays an important role in hybrid methods and that attribute selection should be performed. Since the results of Sequential CBF-CF scores way below the classical models, we therefore focus our discussion on the other two algorithms from now on.

Sensitivity of Attributes To analyze the impact of attributes on attribute-aware CF algorithms, we further partition the ten trials into subsets by varying the number of useful attributes by attribute frequency and χ^2 for each trial. The average sensitivity of attributes from the taxonomy of ten trials

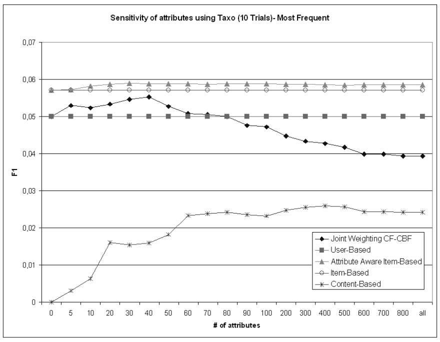

Fig. 3. Vary # of Most Freq. attributes

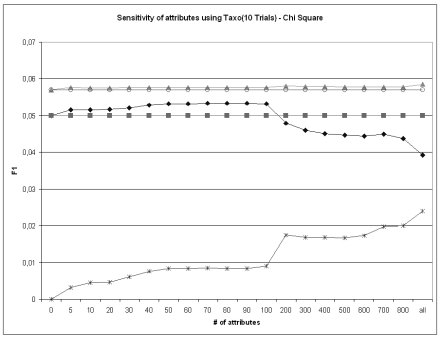

Fig. 4. Vary # of best χ^2 attributes

for each subset are presented in Fig. 3 and Fig. 4.

As shown in both figures, the selection of attributes does affect the quality of topN recommendation. In joint weighting CF-CBF, the quality increases gradually, reaches its peak and decreases dramatically as more irrelevant attributes are appended. In the case of attribute frequency measure, the algorithm reaches its peak at around 40 attributes, whereas in χ^2, the peak is reached in the range of 70-100 attributes. Taking the peaks of both attribute frequency and χ^2 measures, there is an increase of 10.4% and 6.7% respectively compared to its base models. On the other hand, in attribute-aware item-based CF, the quality of attributes has almost no effect on the quality of the recommendations. For attribute frequency, the F1 value quickly meets its peak and maintains rather constant as more irrelevant attributes are added to the algorithms, whereas for the χ^2 measure, the quality reaches the peak when most noise is presented. One of the reasons for these strange results could be due to the value of lambda being set too low as it controls the contribution of attributes to those algorithms.

5 Conclusions and Future Works

Our empirical analysis on state-of-the-art hybrid algorithms shows that the effectiveness of these methods depends on the selection of useful attributes. We have proposed two measures: attribute frequency and chi square. Joint weighting CF-CBF proves to be more effective and provides up to 10.4% gain in F1 than pure CF for movie taxonomy datasets.

As the quality of recommendations varies with the informativeness of the attributes, further studies on other attribute selection measures such as the information gain or the combination of various measures could be the future works.

References

BASILICO, J. and HOFMANN, T. (2004): Unifying collaborative and content-based filtering. In:*21st International Conference on Machine Learning.* Banff, Canada.

BASU, C., HIRSH H., and COHEN, W. (1998): Recommendation as classification: Using social and content-based information in recommendation. In: *1998 Workshop on Recommender Systems.* AAAI Press, Reston, Va. 11-15.

BREESE, J. S., HECKERMAN, D. and KADIE, C. (1998): Empirical analysis of predictive algorithms for collaborative filtering. In:*14th Conference on Uncertainty in Artificial Intelligence (UAI-98).* G. F. Cooper, and S. Moral, Eds. Morgan-Kaufmann, San Francisco, Calif., 43-52.

CLAYPOOL, M., GOKHALE, A. and MIRANDA T. (1999): Combining content-based and collaborative filters in an online newspaper. In: *SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation.*

DELGADO, J., ISHII, N. and URA, T. (1998): Content-based Collaborative Information Filtering: Actively Learning to Classify and Recommend Documents. In: *Cooperative Information Agents II. Learning, Mobility and Electronic Commerce for Information Discovery on the Internet.* Springer-Verlag, Lecture Notes in Artificial Intelligence Series No. 1435.

DESHPANDE, M. and KARYPIS, G. (2004): Item-based top-N recommendation Algorithms. In: *ACM Transactions on Information Systems* **22/1**, *143-177.*

GOLDBERG, D., NICHOLS, D., OKI, B. M. and TERRY, D. (1992): Using collaborative filtering to weave an information tapestry. In: *Commun. ACM* **35**, *61-70.*

GOOD, N., SCHAFER, J.B., KONSTAN, J., BORCHERS, A., SARWAR, B., HERLOCKER, J., and RIEDL, J. (1999): Combining Collaborative Filtering with Personal Agents for Better Recommendations. In: *1999 Conference of the American Association of Artificial Intelligence (AAAI-99), pp 439-446.*

LI, Q. and KIM, M. (2003): An Approach for Combining Content-based and Collaborative Filters. In: *Sixth International Workshop on Information Retrieval with Asian Languages (ACL-2003) pp. 17-24.*

MELVILLE, P., MOONEY, R. J. and NAGARAJAN, R. (2002): Content-boosted Collaborative Filtering. In: *Eighteenth National Conference on Artificial Intelligence(AAAI-2002), pp. 187-192.* Edmonton, Canada.

MOVIELENS (2003): Available at http://www.grouplens.org/data.

PAZZANI, M. J. (1999).: A framework for collaborative, content-based and demographic filtering. In: *Artificial Intelligence Review 13(5-6):393408.*

SARWAR, B. M., KARYPIS, G., KONSTAN, J. A. and RIEDL, J. (2000): Analysis of recommendation algorithms for E-commerce. In: *2nd ACM Conference on Electronic Commerce (EC'00).* ACM, New York. 285-295.

TSO, H. L. K. and SCHMIDT-THIEME, L. (2005): Attribute-Aware Collaborative Filtering. In:*29th Annual Conference of the German Classification Society 2005, Magdeburg, Germany. (to appear)*

ZIEGLER, C., SCHMIDT-THIEME, L. and LAUSEN, G. (2004): Exploiting Semantic Product Descriptions for Recommender Systems. In:*2nd ACM SIGIR Semantic Web and Information Retrieval Workshop (SWIR'04).* Sheffield, UK.