Investigating the use of Association Rules in Improving Recommender Systems

Gavin Shaw
School of Information Technology
Queensland University of Technology
Brisbane QLD Australia
g4.shaw@student.qut.edu.au

Yue Xu
School of Information Technology
Queensland University of Technology
Brisbane QLD Australia
yue.xu@qut.edu.au

Shlomo Geva
School of Information Technology
Queensland University of Technology
Brisbane QLD Australia
s.geva@qut.edu.au

Abstract *Recommender systems are widely used online to help users find other products, items etc that they may be interested in based on what is known about that user in their profile.*

Often however user profiles may be short on information and thus when there is not sufficient knowledge on a user it is difficult for a recommender system to make quality recommendations. This problem is often referred to as the cold-start problem.

Here we investigate whether association rules can be used as a source of information to expand a user profile and thus avoid this problem, leading to improved recommendations to users. Our pilot study shows that indeed it is possible to use association rules to improve the performance of a recommender system. This we believe can lead to further work in utilising appropriate association rules to lessen the impact of the cold-start problem.

Keywords Information Retrieval, Personalised Documents, Recommender Systems, Association Rules.

1 Introduction

Recommender systems are designed to understand a users interests, learn from them and recommend items (whether they be products, books, movies etc) that will be of interest to the user. This requires them to personalise their recommendations. Recommendation systems usually work most effectively when user profiles are extensive and/or the applicable dataset has a high information density. When the dataset is sparse or user profiles are short, then recommender systems struggle to provide quality recommendations. This is often known as the cold-start problem.

Proceedings of the 14th Australasian Document Computing Symposium, Sydney, Australia, 4 December 2009.
Copyright for this article remains with the authors.

Our proposal here focuses on one aspect of the cold-start problem; short user profiles. When a user (can be a new user) has very few ratings in their profile, recommender systems may fail to provide recommendations that interest the user. Both collaborative based [5] and content based [2] can suffer, due to collaborative systems failing to find similar users and content systems having problems due to lack of being able to obtain the content interests of the user.

We propose expanding a user profile (eg. so it contains more ratings) through the use of association rules derived from the dataset. By doing so we expand profiles based on patterns and associations of items, topics, categories etc (which should give relevant consequents) and thus give more information to a recommender system. This would reduce the effect of the cold-start problem and result in better quality recommendations earlier on.

2 Related Work

Much work has been done in the area of recommender systems. Work focusing on solving the cold-start problem includes collaborative & content hybrids [1] [5], ontology based systems [3] and taxonomy driven recommender systems [6] [7]. However, all of these proposals have drawbacks. The hybrid systems can lack novelty, resulting in recommendations that are excessively content centric [7]. The onotolgy based system requires a well defined ontology to be created, something that can be difficult and would limit the system to what is defined/covered within the ontology. Taxonomy based systems work better, but still have low performance. Also the HTR system proposed in [6] performs only marginally better than the TPR system proposed in [7], although it is more time efficient. The taxonomy based approach in [7] does have the advantage of being able to be applied to many domains.

Work has also focused on the other cold-start problem, when a new item is introduced and recommendations are required, but no one has yet rated that item [5]. For this cold-start problem collaborative systems can not help, but content based systems can [5]. We focus on the cold-start problem of new users, rather than new items.

3 Proposed Approach - Using Association Rules to Expand User Profiles

Here we outline our proposed approach and investigation into solving the cold-start problem in recommender systems.

3.1 Background

In its simplest form, a recommender system takes what it knows about a particular user (perhaps a profile) and attempts to predict what that given user would be interested in and recommend those items to the user. Information can take the form of what a user has rated, what other users with similar tastes have rated, the content of the rating (if any) and the content of the item(s). In the case of TPR [7] it uses a given user's rating history of items and the user's implicit taxonomic preferences to determine new items that will interest the user. Both a user's history and implicit taxonomic preferences can be easily obtained from the dataset without any extra effort on the users behalf.

The drawback to this is when a user has a small number of ratings (what we refer to as a 'short profile') it becomes difficult to effectively determine the users implicit taxonomic preferences. Thus recommendation quality suffers.

3.2 Expanding User Profiles

sually at the heart of every recommender system are the user profiles. It is from this that recommenders base their work. In order to improve the quality of recommendations being made to users with short profiles, we propose to use association rules to expand the number of entries in their profiles.

We have a set of users $U = \{u_1, u_2, ..., u_n\}$ and a set of items $I = \{i_1, i_2, ..., i_m\}$. Every user $u \in U$ has a set of rated items $R(u) \subseteq I$ whereby if an item *i* is in the set *R* then user *u* has rated it. We also have a taxonomy *T* containing topics (or categories) *t* in a multi-level structure, where each topic has one parent or supertopic, but may have many children or subtopics. Thus the taxonomy can be visualised as a tree. Each path from the root to a leaf is called a descriptor which is an ordered list consisting of the topics on the path. For any item, it may have more than one descriptor. Let $D = \{d_1, d_2, ..., d_o\}$ be the set of all descriptors, for an item $i \in I$, its descriptors can be represented as $\{d_1(i), d_2(i), ...\}$ which is a subset of *D*. All of this information can be used to expand existing short profiles.

Firstly, using the set of users *U* and the taxonomy *T* we can build a transactional dataset where each user *u* is a 'transaction' and all the topics *t* in the taxonomy make up the datasets attributes. Then we populate the dataset using the set of users *U*, the set of rated items *R* and the set of taxonomic descriptors *D*. This is done by determining the items *i* rated by the user *u* and their positions within the taxonomy. Each item will correspond to one or more paths through the taxonomy from the root to a leaf. That is, the item may have one or more descriptors. For a user u_x and $a \in R(U_x)$, using the descriptors $\{d_1(a), d_2(a), ...\}$, we place a positive value '1' in the user's transaction at each topic involved in these descriptors. All other topics in the transaction will be marked with a negative value '0'. From this we can construct a transactional dataset that shows users' interests in topics, not items.

Second, we then mine the transactional dataset for frequent patterns and derive association rules from these patterns. The frequent patterns and rules will not come from just one taxonomy level, but rather mutiple levels and will also include cross-level patterns and rules. This will give us association rules between topics that interest users. These rules allow us to discover topics that frequently appear together as part of a user's interest. This rule set will then be used to expand user profiles to solve the cold-start problem.

Next, we create the user profiles that will be needed by the recommender system. For our investigation we use the TPR system first proposed in [7]. In order to achieve this we will use the set of users *U*, the set of rated items *R*, the taxonomy *T* and the set of descriptors *D* to create a set of user profiles $P = \{p_1, p_2, ..., p_n\}$. Here for each user $u \in U$ we determine the leaf topics that correspond to each item $i \in R(u_x)$ through the use of the descriptor(s) *d(i)*. These leaf topics are then added to the profile $p(u_x)$ so that $p(u_x)$ contains a list of leaf topics for which user u_x has rated at least one item *i* in each leaf topic *t*. The set of user profiles *P* is known as the user taxonomy profiles and is used by the TPR approach to perform recommendations. This set of profiles *P* will serve as our baseline.

Finally, we expand the user profiles. For this we take the set of user profile *P* and the association rule set we derived in the second step. For each user profile $p(u_x)$ we extract all of the topics *t* within and generate a list of all the combinations possible from the group of topics. Each combination represents a possible antecedent of an association rule. We take each combination and search the set of association rules for any rules that have that exact set of topics as its antecedent. If such a rule exists we can then take its consequent and the topics within and add them to the profile $p(u_x)$. Thus this generates a set of expanded user profiles which we show in our experiments have the potential to improve recommendations over profiles that have not been expanded.

3.3 Imposing Restrictions on User Profile Expansion

We have outlined our proposal for using association rules to expand user profiles in order to improve recommender system quality. However, it is possible that we may want to place limitations on the expansion of user profiles.

1. Restrict the expansion to short profiles. The idea behind this proposal is to expand users who have very few ratings and thus suffer from the cold-start problem. Users with many ratings do not have this problem. Thus a restriction should be imposed on how many topics can be in the user profile *p* before there is too many to warrent expansion. This limit would be dependent on several factors.

2. Restrict the number of rules used when expanding a user profile. It is entirely possible that when deriving the association rules from the transactional dataset that a large list may be generated. It is also possible that from this, when expanding a user profile that a large number of rules and their consequents will be considered for inclusion in the expanded user profile. This may lead to poorer performance as many more topics are added and more items from a wider selection become recommended. Therefore it may be benefical to limit the number of association rules that are used in expansion to those that have the highest support, confidence or other appropriate interestingness measure.

4 Experiments and Evaluation

Here we outline the pilot experiment we undertook to study the value of our proposal to use association rules in expanding user profiles to improve recommendation quality.

4.1 Evaluation Metrics

In order to evaluate the performance of the baseline set of profiles and the expanded set of profiles we follow the same approach detailed in [6]. The past ratings of each user $u \in U$ is divided into a training component and a test component. For the experiments, the recommender system will recommend a list of *n* items for user u_i based on the training set. The recommendation list will be evaluated against the test set. For our experiments we use exactly the same training and test sets as used in [6].

In our work we use precision, recall and F1-measure to determine the overall performance of the recommender system. This allows us to compare the standard approach against our proposal of using association rules for user profile expansion.

4.2 Dataset

For this investigation we use the BookCrossing dataset (obtained from http://www.informatik.uni-freiburg.de/ cziegler/BX/) which contains users, books and the ratings given to those books by the user. The taxonomy tree and descriptors are orginally sourced from Amazon.com and are exactly the same as those used in [6]. From this we build a transactional dataset that contains 92,005 users (transactions) and 12,147 topics from the taxonomy. The dataset is populated using the descriptors that belong to 270,868 unique books. This dataset is then mined to derive the association rules from it.

From the BookCrossing dataset we also build the base set of user profiles *P*. This set of profiles contains 85,415 distinct users with a total of 10,662 leaf topics contained in the taxonomy. As already mentioned the ratings for each user are divided into a training set and a test set. The set of user profiles *P* is based on the training set. The average number of leaf topics in a user profile is 27.08 and the highest number of leaf topics in a given user profile is 3,173 leaf topics. This set of user profiles will serve as the baseline in our experiments and is also the set that will be expanded using the derived association rules.

4.3 Experiment Results

To validate our proposal we conducted a series of experiments to see whether using association rules to expand user profiles improves recommendation quality. From the transactional dataset we set the minimum confidence threshold to 50% and are able to derive 37,827 association rules using the MinMax rule mining algorithm [4]. We then go through the user profiles in the training set and for any profile $p \in P(train)$ that has 5 or less topics listed we attempt to expand using the association rules. This yields a total of 15,912 user profiles which we consider to be short profiles. After attempting profile expansion we then make up to 10 recommendations for these 15,912 users and measure the overall performance of the recommender system. We compare our proposed approach (involving the expanded profiles) against the baseline of the same 15,912 user profiles with no expansion. All experiments use the TPR recommender system first presented in [7].

As shown in Table 1 the baseline set of user profiles (which there is no profile expansion) scores only 0.00619, 0.0571 and 0.0112 for precision, recall and F1-measure respectively. When using expanded profiles we manage to achieve up to 0.00815, 0.0754 and 0.0147 for precision, recall and F1-measure. This is an improvement of approximately 31.5% over the baseline. This level of improvement was achieved when we used the top 5 rules (ranked by their confidence score) to expand user profiles. Table 1 also shows that the performance of our proposed approach improves as more rules are used in expanding a user's profile. However, the improvement is between using the top 2

Table 1: Experimental results for TPR using the short user profiles.

Approach	Precision	%	Recall	%	F1-Measure	%
Baseline (No Rules)	0.00619		0.0571		0.0112	
Expanded (1 Top Rule)	0.00649	4.77%	0.0595	4.28%	0.0117	4.72%
Expanded (2 Top Rules)	0.00714	15.21%	0.0655	14.66%	0.0128	15.16%
Expanded (3 Top Rules)	0.00732	18.15%	0.0672	17.77%	0.0132	18.12%
Expanded (4 Top Rules)	0.00792	27.79%	0.0729	27.75%	0.0143	27.79%
Expanded (5 Top Rules)	0.00815	31.54%	0.0749	31.22%	0.0147	31.51%

or top 3 rules is small. A similar situation also occurs between the top 4 and top 5 rules.

This experiment shows that the overall performance of the recommender system can be improved through the use of our proposed approach, with a 30+% improvement being achieved, which we believe supports our proposal. The efficiency of the recommender is not negatively impacted, as while our expanded profiles take longer to make recommendations for, the time taken is inline with that needed to process a profile with a similar number of topics without profile expansion.

We also conducted a second smaller experiment. We took the 15,912 user profiles that had been deemed to be short and while attempting to expand them, we determined which profiles were actually expanded. From the 15,912 short profiles we were able to expand 11,273 profiles. We then evaluated the recommender system on just these profiles. As Table 2 shows the use of the top 5 association rules to expand these user profiles resulted in an improvement in the performance of the recommender system of 39.7% over the baseline. This experiment shows that for users whose profile can be expanded, noticable improvements in recommendation performance are achievable. Thus it appears that association rules can make a difference in recommender system performance.

Table 2: Experimental results for TPR using only the short user profiles that were sucessfully expanded.

Approach	F1-Measure	%
Baseline (No Rules)	0.0113	
Expanded (5 Top Rules)	0.0158	39.74%

5 Conclusion and Future Work

In this paper we proposed the idea of using association rules to expand user profiles in order to improve recommendations. We outline an approach whereby the rules can be discovered and used, increasing the number of topics in a user profile that only has a few existing ratings. Our experiment shows that the proposed approach can improve the performance of a recommender system under the cold-start problem. This approach allows a user profile to obtain more topic information without extra input from a user and allows a new user to get better recommendations faster.

Further work includes discovering if there is a better measure to rank the rules so that the rules selected are the best for expanding the user's profile. Also more investigation into how many of the top rules to use needs to be undertaken. This would help determine if it is possible to use too many rules during profile expansion such that recommender performance is degraded. Finally, with the issue of redundant rules, this application could be used to help confirm that rules removed as redundant do not cause information loss. This could be done by comparing the performance of a rule set containing redundant rules against one that does not.

Acknowledgements Computational resources and services used in this work were provided by the HPC and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

References

[1] R. Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modelling and User-Adapted Interaction*, Volume 12, pages 331–370, 2002.

[2] P. Melville, R. J. Mooney and R. Nagarajan. Content-Boosted Collaborative Filtering for Improved Recommendations. In *18th National Conference on Artificial Intelligence (AAAI'02)*, pages 187–192, Edmonton, Canada, July 2002.

[3] S. E. Middleton, H. Alani, N. R. Shadbolt and D. C. De Roure. Exploiting Synergy Between Ontologies and Recommender Systems. In *The Semantic Web Workshop, World Wide Web Conference (WWW'02)*, pages 41–50, Hawaii, USA, May 2002.

[4] N. Pasquier, R. Taouil, Y. Bastide and G. Stumme. Generating a Condensed Representation for Association Rules. *Journal of Intelligent Information Systems*, Volume 24, pages 29–60, 2005.

[5] A. I. Schein, A. Popescul, L. H. Ungar and M. Pennock. Methods and Metrics for Cold-Start Recommendations. In *25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'02)*, pages 253–260, Tampere, Finland, August 2002.

[6] L.-T. Weng, Y. Xu, Y. Li and R. Nayak. Exploiting Item Taxonomy for Solving Cold-start Problem in Recommendation Making. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI'08)*, pages 113–120, Dayton, Ohio, USA, November 2008.

[7] C.-N. Ziegler, G. Lausen and L. Schmidt-Thieme. Taxonomy-driven Computation of Product Recommendations. In *International Conference on Information and Knowledge Management (CIKM'04)*, pages 406–415, Washington D.C., USA, November 2004.