Recommender System based on Multi Datasets

B.Jayanthi
Associate professor
Kongu Arts & Science College
Erode, Tamilnadu, India.

K.Duraisamy, PhD
Dean
KSR College of Technology
Thiruchengode, Tamilnadu

ABSTRACT
Recommender systems are one of the tools designed to help users deal with the information explosion by giving information recommendation according to their information needs. The cold-start problem refers to the situations where insufficient initial information of data sources for recommendations to make suggestions to users. When a new user extends into the system initially nothing is known about their preference and this need to be discovered. The process of how to get about it in the quickest and most accurate way is a challenge. This paper were designed in two phases, an off-line phase, where non-redundant multi-level and cross-level association rules and rare-item association rules are built and an on-line phase, where the rules are applied to real situations providing recommendations to customers to solve cold-start problem.

Keywords
Recommender System, Cold-Start Problem, Rare-Item, Association Rule, Non-Redundant.

1. INTRODUCTION
The Web is growing rapidly. More and more information is being made available electronically. Users are often overwhelmed by the huge amount of information and are faced with the great challenge of finding the most relevant pieces of information in a short amount of time. Recommender systems are one of the tools designed to help users deal with the information explosion by giving information recommendations according to their information needs. Recommender systems look at patterns/items that occur together frequently (association) in order to make a recommendation to a person. It should therefore possible to bring them together. Recommender system predicts user's preference to suggest items. When a new user is entered into the system, initially nothing is known about their preferences and thus they need to be discovered. This problem is known as cold-start problem. Moreover, many of the recommender systems are developed using collaborative filtering which is heavily based on users rating data which is very limited as the products are not often purchased by the users during their lifetime, and users are not able to provide ratings for products that they never use. Thus the commonly used recommendation techniques are not suitable for recommending infrequently purchased products.

We have proposed a recommender system which utilizes both frequent and infrequent / rare-item patterns association rules. In this paper we propose a novel method to generate non-redundant multi-level and cross-level rules from multi-datasets. These associations reflect users preferences towards products. It is expected that with the use of multi-level and cross-level association rules and rare-item association rules are used to improve the recommender system. This problem motivates to develop algorithms and techniques to mine frequent patterns/Itemsets, generating non-redundant multi-level and cross-level association rules and effectively using them to improve recommended systems.

2. RELATED WORK
Recently, some researchers have proposed recommender system which utilizes association rules. Collaborative Filtering [4] among others is a promising technique in recommender systems. It maintains data about users preferences, based on which similarities between the users are computed. It then recommends to a target user liked by other, similar users. Collaborative approaches suffer from a problem with sparseness, especially when the number of ratings or preferences is small. Association may be used to reduce this problem. The author [5] proposed a model-based approach that uses multi-level association rules. Association rules are used to compute/determine the preferences for items. More recently an author [6] proposed a collaborative based system that uses association rules to solve the cold-start problem. In their proposed approach fuzzy cross-level rules are mined and used to enhance the recommender with the aim of helping cold-start items when there are no recommendations already made for an item. The authors in [7] proposed an item based approach. It looks at the items users have rated and determines the similarity between a pair of items and the most similar items are then recommended.

The authors in [8] developed an approach to mine both positive and negative multi-level association rules from a dataset. Negative association rules are those infrequent patterns that are in the dataset, which may contain information of interest depending on the user and application. The author in [9] proposed an idea to deal with cold-start problem. He suggests learning about the users preferences that system needs to make accurate recommendations. The authors in [10] proposed hybrid recommender system. They used item taxonomic information, user taxonomic information to improve cold-start situation.

The author [1, 2, and 3] proposed a recommender systems based on using association rules by expanding short user profile to solve cold-start problem. Limitations are imposed on expansion and restrictions to the number of rules to expand a profile. The proposed work will look at using non-redundant multi-level and cross-level association rules and rare-item association rules to solve the cold-start problem.

3. APPROACHES
This study has investigated and developed a novel method to discover an automated approach for mining non-redundant multi-level and cross-level association rules and rare-item

association rules and effectively utilizing them to improve recommender system, in order to solve cold-start problem.

3.1 Generating non-redundant multi-level and cross-level association rules

The pattern of generating non-redundant multi-level and cross-level rules association rule could be decomposed into two sub problems. First, mining of frequent closed itemsets/patterns and generators. Second, multi-level and cross-level association rules are generated from frequent closed itemsets and generators.

3.2 Rule Selection

In rule selection, the switching technique was used; the system switch between association rule based on frequent item and another association rule based on infrequent item to overcome the cold-start problem. Recommendation using association rules is to predict preferences of item k when the user preferred an item i .The system search the set of multi-level and cross-level association rules, for any rules that have that exact set of topics as its antecedent. If such a rule exists, consequent sets from association rules whose antecedent matches is taken. The top-N rules with highest confidence are selected for recommendation. If a such a rule does not exists, the system switch over to infrequent association rules and Search for any rules, which have that exact set of topics as its antecedent and its consequents are taken and recommended.

3.3 Rare-Item Association Rule Mining

In recommendation using association rules, if the amount of available preference information is small, then the number of association rules would also be small. In this case, it is impossible to predict preferences for most of the infrequently purchased items and performance of recommendations becomes very poor. For those, infrequently purchased items are the lack of data about user's interests, the hidden information stored in transaction database could considered as another important resource for making recommendation. To overcome this problem, a Rare-Item/infrequent itemset association rules are generated for recommendations. In order to generate a rule, previously during pruning process infrequent itemsets are stored in a Rare-Item table. Rare-Item is indexed with its transaction lists so as to find maximally itemsets/patterns to generate association rules.

3.4 Rare-Item Indexing

In this work, the Rare-Items are kept in the Table 1 . The Rare-Item Index table contains the following attributes: (Item, TIDList). For every unique higher level item, item id is listed with its list of all the transactions. This indexing reduces the number of scanning in the databases and reduces the running time.

3.5 Properties of Rare-Item association Rule mining

A proposed method for Rare-Item association rule assumes that a transaction that supports an item also support its attributes. The following properties are used in the proposed work to mine Rare-Item association rule.
Property 1:If an Item is frequent, its attributes are also considered frequent.
Property 2: If an item is infrequent, it is still possible for all or some of its attributes to be frequent.
Property 3: Rare-Item is the antecedent part of the rule and the maximally itemsets generated from the frequent/infrequent attributes is the consequent part of the rule.

Table.1: Rare-Item Index Table

S.No	Item	TID List
1	1-1-2	3
2	1-1-3	6
3	1-2-1	1
4	1-2-2	3,5
5	3-2-3	2,6,8
6	4-1-1	3,8
7	4-1-3	5
8	5-2-4	6,8
9	7-1-3	8

3.6 Rule mining

Rare-Item is retrieved with their transaction lists i.e., Rare-Item Index Table. For every Rare-Item, Let TID-Lists be the set all transactions. Calculate the maximally itemsets from TID-lists of every Rare-Item. Association rule is generated with Rare-Item as antecedent and maximally itemsets with higher level item as consequent and it is placed in Rare-Item association rule table. Since, the Rare-Item is focused on infrequent itemsets as well as to improve recommender systems; instead of support-confidence framework, maximally support is used.

The following table shows the rare-item association rule.

Table 2: Rare-item association rule

S.No	Rare-ItemAssociation Rules
1	1-1-2 → 2-*-*,4-*-*
2	1-1-3→3-*-*,5-*-*
3	1-2-1→ 2-*-*
4	1-2-2→2-*-*,4-*-*
5	3-2-3→5-*-*
6	4-1-1→1-*-*,3-*-*,5-*-*,7-*-*
7	4-1-3→1**,2**
8	5-2-4→3-*-*
9	7-1-3→3-*-*,4-*-*,5-*-*

3.7 Algorithm RECITEM

Input: (1) Transaction Table – TID
(2) Association rule table
(3) Rare-Item Association rule Table
(4) Preference Item from Target User
Output: Top N Recommendation rules
Steps
1. Get the preference Item from the target user
2. Find the rules whose antecedent matches with the user preferred item.
3. If such a rule exists
Select the corresponding consequents and Go to Step4

Else

Find the item in the Rare-item association rule table and make recommendations

4. Sort the rules in descending order based on the Confidence values

5. Select Top-N rules and Recommend.

The steps of the algorithm are explained as follows: In this work, two possible solutions are stated to solve cold start problem. First, when the user preferred item i, then the multi-level and cross-level association rule table was searched for any rules that have that exact set of topics as antecedent. If such a rule exists, then its consequent sets from association rule was taken and if more consequent set matches then it may be beneficial to select the rules with highest confidence. The top-N items are recommended to the target user based on the selected rules. Secondly, if any association rules that do not have exact set of topics as antecedent with user preferred item i, then item is searched with Rare-Item association rule table for the preference item.

3.8 Working of RECITEM

This section shows the example to demonstrate the proposed algorithm to recommend items.

Table 3: Sample Database

TID	Items purchased – GID
1	1-1-1, 1-2-1, 2-1-1, 2-2-1
2	1-1-1, 2-1-1, 2-2-2, 3-2-3
3	1-1-2, 1-2-2, 2-2-1, 4-1-1
4	1-1-1, 1-2-1
5	1-1-1, 1-2-2, 2-1-1, 2-2-1, 4-1-3
6	1-1-3, 3-2-3, 5-2-4
7	1-3-1, 2-3-1
8	3-2-3, 4-1-1, 5-2-4, 7-1-3

The Database in Table3 consists of TID and Items Purchased. TID is the Transaction id number which is unique for every transaction. Items purchased represent the number of items purchased per transaction. Here thirteen items 1-1-1,1-1-2,1-1-3,1-2-1,1-2-2,1-3-1,2-1-1,2-2-1,2-3-1,3-2-3,4-1-1,5-2-4 and 7-1-3 are considered.

When the user preferred item 2-1-* then the multi-level and cross-level association rule table was searched and top-N items are recommended to the target user based on the confidence. When the user preferred item 3-2-3 then the multi-level and cross-level association rule table was searched and such a rule does not exists, then item is searched with Rare-Item association rule table for the preference item, then the corresponding consequents are recommended. Thus, user can obtain recommendations through the frequent itemset based association rule until the system has enough information. If there was not enough information to use frequent itemset based association rule, the system would use rare-item association rule.

Table 4: Association rules and the corresponding confidence values

S.No	Rules	Support	Confidence
1.	2-2-* → 1-*-*	0.571	1.0
2.	1-2-*-→1-1-*	0.571	1.0
3.	2-2-*→1-1-*	0.571	1.0
4.	2-1-*→ 1-*-*,2-2-*	0.428	1.0
5.	2-1-*→ 1-1-*,2-2-*	0.428	1.0
6.	2-2-1→1-1-*,1-2-*	0.428	1.0
7.	2-1-*→1-1-1,2-2-*	0.428	1.0
8	2-2-*,1-1-1→2-1-*	0.428	1.0

Table 5: The rare-item association rule

S.No	Rare-Item Association Rules
1	1-1-2 → 2-*-*,4-*-*
2	1-1-3→3-*-*,5-*-*
3	1-2-1→ 2-*-*
4	1-2-2→2-*-*,4-*-*
5	3-2-3→5-*-*
6	4-1-1→1-*-*,3-*-*,5-*-*,7-*-*
7	4-1-3→1-*-*,2-*-*
8	5-2-4→3-*-*
9	7-1-3→3-*-*,4-*-*,5-*-*

4. EXPERIMENTAL EVALUATION

Experiments are conducted on Book Crossing Dataset which contains users, books and ratings given those books by the user. From this a transaction dataset that contains 92,005 transactions were built. This dataset is mined for frequent 1 to k itemset in chapter 4. Non-redundant multi-l and cross-level association rules and rare-item association rules were derived in chapter 5. To validate the proposed work a series of experiments were conducted. Dataset is divided into 80% of training set and 20% of test set and averaged the results. Experiments are performed with two different methods of applying association rules. First method is applying non-redundant multi-level and cross-level association rules. The second method is using Rare-Item association rules.

Recommendation quality is based on accuracy. Accuracy is computed as the ratio of correct prediction. If the user's real expectations are actually recommended – contained within the top-N recommendation list, it's called hit. Recommendation quality is evaluated based on precision, recall, and f-score.

Let the set of items relevant to a user preferred item be denoted as Relevant, and the set of recommended items be denoted as Recommended.

Precision: This is the percentage of recommended items that are relevant to the user preferred item.

Recall: This is the percentage of the items that are relevant to the user preferred item.

F-Score, which is defined as the harmonic mean of recall and precision.

$$\text{F-Score} = \frac{\text{recall x precision}}{(\text{recall + precision})/2}$$

Initially, User expected preference data is kept in test set and user is expected to state preferences, if the recommended top-N items with highest predicted preference are based on user's preferences that represent what they are looking for. It is more likely that better recommendations can be obtained.

Table 6 shows the two sets of association rules are derived from the transactional database with the minimum confidence threshold set to 50% and different minimum support settings for the two set of rules. One set of rules derived using a set of minimum support values set at 10%, 7.5 % and 5% for 3 concept levels and another set of minimum support value set at 10%,8% and 6%.For each set of minimum support values Reliable rules (Reliable ExactRule and Reliable ApproximateRule) with reduced hierarchical redundancy is derived. The cold-start itemsets are randomly selected as C%(C = 5, 10, 20) and Top-10 Recommendations are generated.

Table 6: Rulesets

S.No	Ruleset ID	Minimum Support Values	Rule Extraction approach	Ruleset Size
1.	RS1	10%l,7.5%,5%	Reliable Rule with Modified CLOSE+	36,585
2.	RS2	10%,8%,6%	Reliable Rule with Modified CLOSE+	8,657

Table 7: Recommendation Accuracy for Frequent itemset

Ruleset ID	Metric	C = 5	C = 10
RS1	Precision	0.099	0.096
	Recall	0.101	0.098
	F-Score	0.099	0.097
RS2	Precision	0.094	0.096
	Recall	0.099	0.098
	F-Score	0.096	0.097

Table 7 and 8 reports the algorithm RECITEM recommendation accuracy, which declines in general as infrequent itemset increases. This suggests that generating recommendations becomes more difficult when there are more cold-start items. This is due to decreasing number of user preferences for an itemset. This algorithm works better with both frequent and infrequent itemset based association rules.

Table 8: Recommendation Accuracy for rare itemset

Ruleset ID	Metric	C = 10	C = 20
RS1	Precision	0.074	0.072
	Recall	0.101	0.061
	F-Score	0.085	0.066
RS2	Precision	0.075	0.072
	Recall	0.099	0.064
	F-Score	0.085	0.068

The recommendation quality of non-redundant frequent itemset based association rule sets of RS1 and RS2 are shown in the following figures 1 and 2.

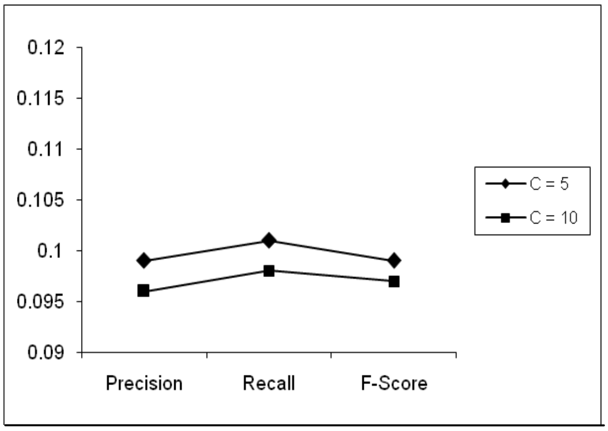

Figure 1: Recommendation quality for rule set RS1

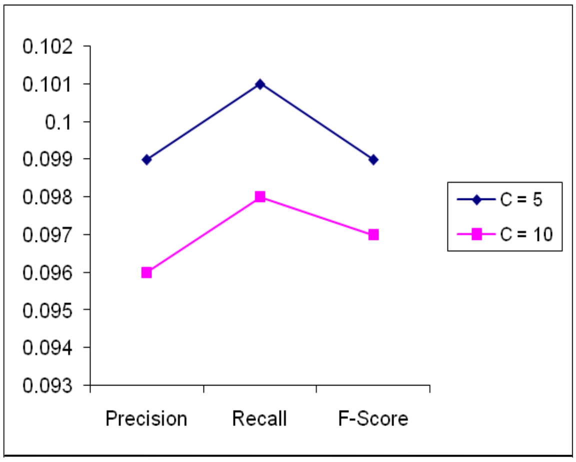

Figure 2: Recommendation quality for rule set RS1

The recommendation quality of Rare-itemset based association rule sets of RS1 and RS2 are shown in the following figures 3 and 4.

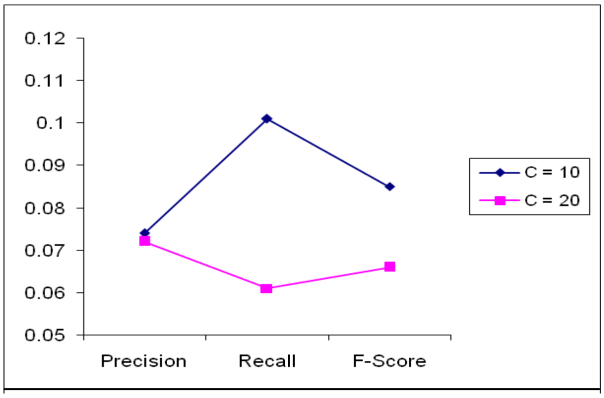

Figure 3: Recommendation quality for Rare-Itemset rule set RS1

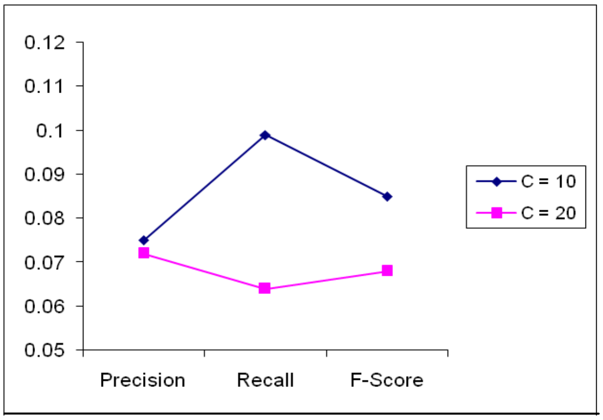

Figure 4: Recommendation quality for Rare-Itemset rule set RS2

5. CONCLUSION

The proposed method presented in this paper use Rare-item association rules in order to improve recommender system performance when faced with the cold-start problem. Experiments are performed on real datasets. The experiment shows the performance of the proposed RECITEM algorithm in recommending the user preferences, what they look for. Experimental results reveal the efficiency of the proposed algorithm.

6. REFERENCES

[1] Gavin Shaw, Yue Xu, and Shlomo Geva, Investigating the use of Association Rules in Improving Recommender Systems, Proc. 14th Australasian Document Computing Symposium, Sydney, Australia, pages 106 – 109, 2009.

[2] Gavin Shaw, Yue Xu, and Shlomo Geva, Using Association Rules to solve the Cold-Start problemin Recommender Systems, In Proceedings of 14th Pacific-Asia Conference on Knowledge Discovery and Data, Hyderabad, India, 2010.

[3] Gavin Shaw, Discovery & Effective use of Quality Association Rules in Multi-Level Datasets, Ph.D Thesis, Queensland University of Technology, Australia, 2010.

[4] M. Kaya, and R. Alhaji, Mining Multi-Cross-Level fuzzy weighted association rules, Proc. International IEEE Conference on Intelligent Systems, June, 2004.

[5] C. Kim, & J. Kim, A Recommendation Algorithm using Multi-Level Association Rules, Proc. IEEE/WIC International Conference on Web Intelligence, October, 2003.

[6] C. W. Leung, S. C. Chan, and F. Chung, An Empirical study of a cross-level association Rule Mining approach to Cold-Start Recommendations, Knowledge Based Systems, pages 512 – 529, 2008.

[7] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, Item-Based Collaborative Filtering Recommendation Algorithms, in Proceedings of 10th International World Wide Web Conference, Hong Kong, May, 2001.

[8] L. K. Sharma, O. P. Vyas, U. S. Tiwary, and R. Vyas, A Novel Approach of Multilevel Positive and Negative Association Rule Mining for Spatial Databases. In P. Perner & A. Imiya (Eds.), Machine Learning and Data Mining in Pattern Recognition Springer Berlin / Heidelberg. Vol. 3587, pages 620 - 629, 2005.

[9] Y. Z. Wei, A Market-Based Approach to Recommender Systems. University of Southampton, 2005.

[10] L. –T. Weng, Y. Xu, Y. Li, R. Nayak, Exploiting Item Taxonomy for Solving Cold- Start Problem in Recommendation Making, in proceedings of IEEE International Conference on Tools with Artificial Intelligence, 2008.