Implementation of Movie Recommender System using Association Rule Mining

Wai Wai Khine, Thinn Lai Lai Thein
Computer University, Sittwe
waiwaikhine15@gmail.com

Abstract

Collaborative recommender systems allow personalization for e-commerce by exploiting similarities and dissimilarities among users' preferences. This paper presents an approach to using data mining for e-commerce. It applies association rule mining to collaborative recommender systems, which recommend articles to a user on the basis of other users' ratings for these articles as well as the similarities between this user's and other users' tastes. This approach makes recommendations by exploring associations between users, associations between articles, and a combination of the two. It is found that association rules are quite appropriate for collaborative recommendation domains and that better performance can be achieved that is comparable to current state of the art in recommender systems research.

1. Introduction

In observing the user, there are various classes of users watch to specific movie class. There are a variety of movies already. There is a difficulty of searching which kinds of movie he/she likes. Recommending the users based on his/her attributes, helps the user's listening patterns and automatically personalizes the interface presentation so as to aid the use in making listening or downloading.

Today, Users have more choices than ever. They are more aware of the possibilities and demanding of personal attention. It is now becoming increasingly important for a company to build a strong relationship with its users. This is where personalization technology steps in. Users can be classified into "classes" based on past behavior as well as predictions of future behavior. Each of these "classes" can then be treated differently, based on marketing needs and behavior, providing the individual attention and offers that match users' real interests.

Recommender systems try to recommend articles of potential interest to a user with respect to the user's individual references. Such recommender systems are the focus of current interest in part because of their importance for electronic commerce. Currently, online recommendation services span the areas of book, music, movie, web page and restaurant recommendations, demonstrating the wide range of application domains of existing recommender systems.

In content-based filtering system, a user profile represents the content descriptions of items in which that user has previously expressed interest. The primary drawbacks of content based filtering systems are their tendency to overspecialize the item selection since profiles are solely based on the user's previous rating of items.

Collaborative filtering system has tried to address some of the shortcomings of other approaches mentioned above. The major drawback of collaborative based filtering is that similarity algorithm requires that the neighborhood formation phase be performed as an online process. As the numbers of users and items increase, this approach may lead to unacceptable latency for providing recommendations or dynamic content during user interaction.

Therefore, this system use model based personalization system because of above drawbacks. Association rules have been used with success in other domains.

2. Related Work

Almost all recommender systems fall into two categories: content-based recommendation and collaborative recommendation. As mentioned in [2], content-based recommendation tries to recommend articles similar to those articles the user has liked, whereas collaborative recommendation tries to find some users who share similar tastes with the given user and recommends articles they like to that user. Content-based recommendation and collaborative recommendation both have their own advantages and drawbacks. But collaborative recommendation is more popular than content-based recommendation, mainly because in many domains (such as music, restaurants) it is hard to extract useful features from

articles, which is generally a step required for content-based recommendation.

Although there seems to be an increasing commercial demand for collaborative recommendation techniques, both the number of available published techniques and the information about their performance are quite limited. Hence it is of great importance to explore more techniques for this domain. Through our analysis and experiments, it has been found that association rules are quite appropriate for this task and they can achieve good performance.

3. Recommender System

Recommender systems are systems that provide the product information and suggestion, help customers decide what products to buy and assist users complete the shopping process through the simulation process. The products can be recommended based on the top overall sellers on a site, based on demographics of the customer, or based on an analysis of the past buying behavior of the customer as a prediction for future buying behavior.

Web-based recommender systems have become a very popular tool that enables to deliver personalized information for every Internet user. The ability of a recommender system to tailor its output to a particular user implies that it must be able to infer what that user requires based on previous or current interactions with that user or other similar users. They connect users with items to "consume" (purchase, view, listen to, etc.) by associating the content of recommended items or the opinions of other individuals with the consuming user's actions or opinions. Such systems have become powerful tools in domains from electronic commerce to digital libraries and knowledge management. For example, a consumer of just about any major online retailer who expresses an interest in an item – either through viewing a product description or by placing the item in his "shopping cart" – will likely receive recommendations for additional products. These products can be recommended based on the top overall sellers on a site, on the demographics of the consumer, or on an analysis of the past buying behavior of the consumer as a prediction for future buying behavior.

4. Association Rule Mining

Association rules have been widely used in web mining. They have been used to mine path traversal patterns and to facilitate the best design. It is a good framework to mine association rules in navigation history for recommending web pages. Traditional association rule mining algorithms are not good enough for recommender systems.

4.1 Association Rule Implementation

Given a set of transactions, where each transaction is a set of items, an association rule is a rule of the form X ==> Y, where X and Y are sets of items. The meaning of this rule is that the presence of X in a transaction implies the presence of Y in the same transaction. X and Y are respectively called the body and the head of the rule. Each rule has two measures: confidence and support. The confidence of the rule is the percentage of transactions that contain Y among transactions that contain X; The support of the rule is the percentage of transactions that contain both X and Y among all transactions in the input data set. In other words, the confidence of a rule measures the degree of the correlation between itemsets, while the support of a rule measures the significance of the correlation between itemsets.

It is assumed to have a database of transactions as listed in Table 1, for association rule "{A} ==> {C}", the confidence of the rule is 66%, and the support of the rule is 50%. Table 1 presents the sample transaction for purchased items in sales vouchers. There could be any number of items present in the body and in the head of a rule. A user could also specify some rule constraints, for example, he/she might only be interested in finding rules containing certain items.

The traditional association rule mining problem definition is: given a set of transactions, where each transaction is a set of items, and a user-specified minimum support and minimum confidence, the problem of mining association rules is to find all association rules that are above the user-specified minimum support and minimum confidence.

Table 1: Sample Transaction

Transaction Id	Purchased Items
1	{A, B, C}
2	{A, D}
3	{A, C}
4	{B, E, F}

4.2 Process of Association Rule

A set of items is called an itemset. The support of an itemset is the percentage of transactions that contain this itemset among all transactions. An itemset is frequent if its support is greater than the user-specified minimum support. The problem of

discovering association rules could be decomposed into two subproblems:

- Find all frequent itemsets.
- Generate association rules from frequent itemsets: for example, if {a, b, c, d} and {a, b} are frequent itemsets, then compute the ratio:

  confidence = support{a, b, c, d} / support {a, b}

  If confidence is not less than the user-specified minimum confidence, then "{a, b} ==> {c, d}" is one desired association rule. This rule satisfies the minimum support constraint because {a, b, c, d} is a frequent itemset.

The second part of the process described above is relatively straightforward. But the process of discovering frequent itemsets is computationally expensive, usually requiring multiple passes over the whole database.

5. Proposed System

This paper presents a framework for mining association rules for recommender systems. Rules like "90% of users who like article A and article B also like article C, 30% of all users like all of them and 90% of articles liked by user A and user B are also liked by user C, 30% of all articles are liked by the three of them" are useful for recommendation systems.

It implements the process of making recommendations by using user associations, article associations and a combination of the two. Article associations represent relationships among articles and user associations represent relationships among users. Article associations and user associations can be explored on two levels (like and dislike) by using extensions of the basic association rules. Figure 1 presents the proposed system overview.

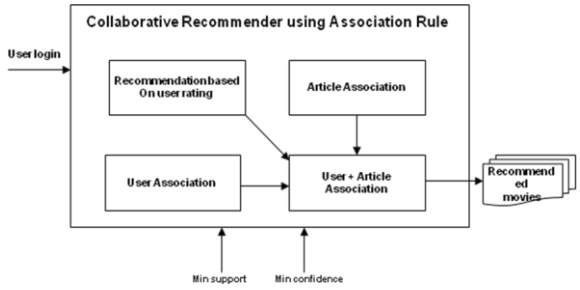

Figure 1: Proposed System overview

5.1 Recommendation Process

Process of making recommendations is performed by using user associations, article associations and user to article association. Article associations represent relationships among articles and user associations represent relationships among users. Article associations and user associations can be explored on two levels (like and dislike) by using extensions of the basic association rules.

User similarity can be defined in terms of behavior, interests, preferences etc that can be modeled.

- Pattern discovery is performed over the transaction data.
- The contents of the user's transaction are matched against the discovered patterns.
- The matches are ranked.
- The objects associated with the best matches are ranked excluding objects already seen.
- The objects with the higher-most rank are shown to the user.

It matches left-hand side of rules with the active user session and recommend items in the rule's consequent. It is essential to store patterns in efficient data structures, providing good recommendation accuracy. High support thresholds lead to low coverage and may eliminate important, but infrequent items from consideration. Low support thresholds result in very large model sizes and computationally expensive pattern discovery phase.

5.2 Process flow of the System

This system is the implementation of movie recommender system using adaptive association rules. It consists of the following steps:

Begin

User Login to the system
Load User profile from database
Find similar user group for calculating user association rules
Recommender set using user association rules
Recommender set using item association rules
Recommender set using user to item association rules
Combine recommender sets and Recommend to the user

End

Figure 2 illustrates the process flow of the system.

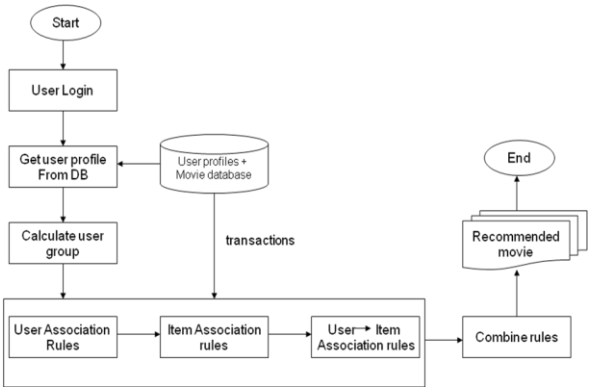

Figure 2: Process flow of the System

5.3 Components of the System

This system consists of three main components:

- Finding user association rules – for example, "90% of articles liked by user A and user B are also liked by user C, 30% of all articles are liked by all of them".
- Finding item association rules – Similar to the user associations, in order to mine like associations among items, each item corresponds to an "item" and each user corresponds to a "transaction".
- Finding user → item association rules – This phase recommend user based on the rules, that contain which user likes which items with support and confidence values.

6. System Implementation

This system is implemented as Web-based movie recommender system and developed using Microsoft Visual Studio 2008 with ASP .Net C#. Movies used in this system are downloaded from IMDB (online movie database).

This system is tested with different rule lengths and different minimum confidences. It calculates the item association, user association and item to user association. With different rule length Table 2 and Figure 3 describe the performance analysis of the system.

Table 2. Performance of different rule length

Rule length	2	4	6	8	10
Accuracy	0.694	0.696	0.695	0.697	0.694
Precision	0.704	0.724	0.733	0.738	0.736
Recall	0.572	0.545	0.529	0.528	0.520

With different rule length, we got the results as shown in Figure 3.

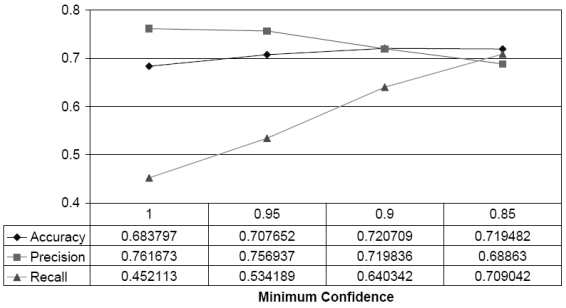

Figure 3: Performance for Different Minimum Confidence

7. Conclusion

This system presents a framework for applying association rules for recommender systems. With the drawbacks of content-based and collaborative recommender systems, association rule mining solves the problems in recommender systems.

It improves the collaborative recommender system by finding the association rules among users, articles and user and articles. Experimental results show that under similar experimental conditions, the performance of user associations is better than that of article associations. It is found that this approach can satisfy the real-time recommendation requirement (especially when user and article associations are combined), and can achieve very good performance, which is at the same level of the current state-of-the-art collaborative recommendation techniques.

8. REFERENCES

[1] J. Breese, D. Heckerman, and C. Kadie. "Empirical analysis of predictive algorithms for collaborative filtering", In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, July 1998.

[2] M. Balabanovic and Y. Shoham, "Content-based, collaborative recommendation", Communications of the ACM, 40(3):66–72, March 1997.

[3] P. H´ajek and T. Havranek, "On generation of Inductive hypotheses", Int. J. Man-Machine Studies, 9:415–438, 1977.

[4] P. H´ajek, I. Havel, and M. Chytil, "The guha method of automatic hypotheses determination". Computing, 1:293–308, 1966. [HHC66]

[5] R. Agrawal, T. Imielinski, and A. Swami. “Mining association rules between sets of items in large databases”. In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 207–216, Washington, D.C., May 1993. ACM.

[6] R.Agrawal and R. Srikant. “Fast algorithms for mining association rules”. In Proc. of the 20th VLDB Conference, pages 487–499, Santiago, Chile, 1994.

[7] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for market basket data”. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 255–264, New York, May 1997. ACM, ACM Press.