An Effective Recommendation Framework for Personal Learning Environments Using Hybrid Recommendation Algorithms

G.Vijayalakshmi, V.Pandiyaraju

PG Scholar, Research Scholar

Department of Information Science and Technology, Anna University, Chennai, India.

Abstract—In Personal Learning Environments, personalized recommendations are used to support the activities of learners and deliver the suitable learning resources to learners. It models the dynamic multi preferences of learners using the multidimensional attributes of resources and learner ratings. By using the data mining technology the cold start and sparsity problems are eliminated. It also increases the diversity of the recommendation list. It has two main modules namely an explicit attribute based recommender and an implicit attribute based recommender system. Based on the explicit multidimensional attributes of resources and historical ratings of the accessed resources the learner preference tree (LPT) built for each learner to express the interests of the learner. The weights of implicit or latent attributes of resources of the learner are considered as chromosomes in a genetic algorithm (GA) and this algorithm optimizes the weights according to the historical ratings in the second module. In nearest neighborhood collaborative filtering (NNCF), the recommendations are generated.

Keywords — Collaborative filtering, learning environment, sparsity, personalized recommender, genetic algorithm, implicit attributes, explicit attributes.

I. INTRODUCTION

In recent years, the web based education has undergone rapid development. With the growth of many online learning systems, the learning resources available in web are heterogeneous and in various media formats. There is a massive increase in product variety, the probability of learners accessing the relevant items is of greater concern and is intensively researched by the Technology Enhanced Learning (TLE) community. In Learning environments to address the information overload and personalization problems the recommender systems have been proposed. The recommender system is a software tool and which provides suggestions to the user to various decision-making processes. It is a software agent that recommends useful and interesting learning resources to the learner by accounting the ratings, preferences and expertise of learners. In the web based education, the greatest challenge does that people prefer differs a lot in their interest and is necessary to provide personal learning environments that can adapt to the levels and interests of learners.

The event, session and recommendation process are the basic key elements of the recommender system. A call to the system provided by an action performed by the user is an event. Every call on a hyperlink generates a new event session is a set of close events provoked by the user u. The sequence of actions that a recommender executes to produce a set of recommendations is called the recommendation process. The system recommends to the user is denoted as an item. A recommendation event contains one or more sessions. The recommendation event is the set of items available to be recommended, for each event the recommendation window was created, a filter for creating and filling a window and guide to wrap and display the items to be recommended.

II. RESEARCH CHALLENGES

With the huge amount of e-learning resources and the digitalization conventional resources, it is difficult for learners to discover the most appropriate resources using a keyword search method. In the e learning environment, sparsity and cold start problems are the most important one. The overspecialized recommendation list, which occurs, is the second important problem in the e-learning environment.

A. Cold Start

The situation in which an item cannot be recommended unless it has been rated by a substantial number of users. It is particularly detrimental to users. To provide reliable and accurate recommendations, a new user has to rate a sufficient number of items before the recommendation algorithm.

B. Sparsity

This problem occurs when rating data are insufficient for identifying similar users (neighbors). But in real time recommender system engaged to evaluate very large data sets, and since each user only rates a small number of items. The number of ratings given by the users is very small in comparison with the total number of (user, item) pairs in the system which leads to sparsity problem.

C. Overspecialized Recommendation

This problem occurs when recommended items are very similar to each other and the recommendation lists are not in diverse. The main reason for overspecialization is inadequate attribute information about users and items. The most vital challenge relates is how to gather the attribute information and use it for modeling the multi preferences of a learner in the resource recommendation. Such challenges are served as potent motivators to identify the suitable learning resources for the users and also find the suitable recommendation techniques.

III. RECOMMENDER SYSTEM APPROACHES

While building an e-learning recommendation
systems there are two main approaches are normally used namely
- Content-based filtering
- Collaborative based filtering
- Hybrid recommendation

A. Content Based Filtering
In this technique, it recommends items similar to the ones that each user liked and evaluated in the past. Based on the matching of their attributes to the user profile the data mining techniques recommend the items to the user. Case-based reasoning mechanisms used to evaluate all the cases in the case base to retrieve the most similar cases. But the performances of case-based reasoning mechanism are similar to that of indexing approaches, so their superior performances are unstable and not guaranteed and it recommends the similar items so it will cause the over specialization recommendation list.

B. Collaborative Filtering
It is one of the most important and useful strategies in the recommender systems. In an e-learning environment, the collaborative filtering approach mainly focused on the correlations among the users having similar interests. Collaborative Filtering can be divided into three categories namely neighbor based collaborative filtering, Model based technique and Demographics approach. The main drawbacks of this approach are the available data are insufficient for identifying similar users which lead to occur sparsity problem and its quality and applicability is limited and to get the accurate recommendation for the given user, it needs to know many user profiles.

C. Hybrid Recommendation
Hybrid recommendation strategies combine several input hybrid sources or several recommendation strategies which provide the better results when compared to use the single strategy alone.

IV. LITERATURE SURVEY
For better understanding purpose, the literature review is given below. Several approaches have been proposed and developed in the past for effective recommendation framework.

Baudisch [1] proposed A new system architecture that supports the formulation of universal queries by joining the table in the sense of a relational database. Walker A et al [2] made a review on collaborative filtering technique and proposed a new system called Altered Vista using the collaborative information filtering approach evaluating the educational effectiveness and the usefulness of the approach. Its limitation is the cold start problem will occur. Khribi M. K, Jemni[5] proposed Automatic Recommendations for e-learning Personalization based on web usage mining techniques and information retrieval provides building automatic recommendation in e-learning. It builds learner and content models in offline module and an online module it recognizes student needs and goals and predict recommendation list. Its limitation is failed to detect learner’s profile, knowledge and preferences automatically in e-learning system.

Khairil Imran Bin Ghauth [4] proposed Building an e-learning Recommendation System Using Vector Space Model and Good Learners Average Rating. It recommends learning materials based on the similarity of the content items and learner’s rating strategy. Its limitation is the performance analysis and benchmarking comparisons are not scheduled.

Enrique Garcia and Sebastian Ventura [6] proposed Architecture for Making Recommends to Courseware Authors Using association rule mining and collaborative filtering. It provides system oriented to find, share and suggest the most appropriate modification to improve the effectiveness of the course. It is failed to group more students, groups, experts and teachers from the other area and obtain more heterogeneous teacher’s profile.

Jesus Bobadilla et al [7] proposed an improving collaborative filtering recommendation system results and performance using genetic algorithms which provides the metric to measure the similarity between users which is applicable in collaborative filtering processes carried out in recommendation system. Its limitation is highly depends on equipment overload and offline process is costly. Heung Nam Kim, Abdul masjid Alkhaldi et al [8] proposed Collaborative user modeling with user generated tags for social recommender systems. It provides a better representation in user interests and achieves better recommendation results in terms of accuracy and ranking by using collaborative approach. Its limitation is the cold start problem will occur.

Subhash K. Shinde et al [10] proposed Hybrid personalized recommendation system using centering-bunching based clustering algorithm. It collects opinions from the user and clustered in offline using CBBC and store in database for future recommendation. The recommendations are generated in online for the active users. Its limitation is the sparsity problem occurs. Wen Li et al [9] proposed Design of a Personalized Learning System Based on Intelligent Agent for E learning. It provides personalized learning environment for the users based on the intelligent agent. Its limitation is the security is not guaranteed.

V. METHOLOGY
In this paper the methodology contains two main modules namely
- Explicit attribute based module (EAB)
- Implicit attribute based module (IAB)

In Fig 5.1it models the user’s preferences and computing the appropriateness degree of items for a target user is the most important task of a recommendation system. The traditional methods are too time consuming and very difficult so use various preference elicitation methods in the current decision making systems. The Matrix Factorization (MF) is used for implicit attribute extraction and Genetic Algorithm (GA) used to discover the explicit attributes.
A. Explicit Attribute Based Module

In Explicit Attribute based module the Learner Preference Tree (LPT) is built for each learner which reflects the learner’s complete spectrum of interests. It takes the resource’s multidimensional attributes and learner’s rating information simultaneously. Based on changes in the ratings of learners for multidimensional attributes of resources the recommendation results can change.

1. Material Modeling

Consider some attributes for the learning material and the ratings of the learner’s accessed materials, which have certain attribute values. The material attributes can be defined as a vector \( M = ((A_1, AW_1), (A_2, AW_2), \ldots, (A_n, AW_n)) \)

Where \( A \) denotes the attribute name of the material and \( AW \) denotes the attribute weight of the material. The attributes of the material is defined as \( M_j = (A_{K1}, A_{K2}, \ldots, A_{Kn}) \) where \( A_{K} \) - attribute keyword of the material \( M_j \).

In this module the learner’s primary subject, secondary subject, educational level and publisher are taken as attributes and the attribute keywords are determined by the experts when a material is first registered in the system for the first time.

2. Frequent Sequences Finding

Web usage mining technique such as prefix span algorithm is used to find the frequent sequential Pattern of an accessed material.

Algorithm:

Prefix span

Input:

The database \( S \) and the minimum support threshold \( \text{min}_\text{support} \).

Output:

The complete set of sequential patterns.

Subroutine:

Prefix Span \((\alpha, L, S|\alpha)\).

Parameters:

\( \alpha \): sequential pattern,
\( L \): the length of \( \alpha \);
\( S|\alpha \): the \( \alpha \)-projected database

Call prefix Span \((<>, 0, S)\).

Method:

Step 1:

Find the length of sequence patterns for the Sequence database \( S \) and consider the minimum support that has been specified.

Step 2:

Divide the search space into the prefixes whose support is greater than the minimum support.

Step 3:

The subsets of sequential patterns can be searched by constructing projected databases of the prefixes supported. Searching is either can be depth first or breadth first search.

Step 4:

Mine projected database repeatedly for sequential patterns until frequent sequences of the database are not found.

3. Interest Learner Modeling

In this Learner Preference Tree is built for each learner. It combines the multidimensional attributes of accessed resources and learner’s rating information as shown in Fig 5.2. It has \( (m+1) \) level starting from 0. The leaf node contains accessed resource ID of the learner and rating. The non-leaf node contains the keyword of the attribute and rating. Each accessed resource corresponds to the unique path from the root to the leaf node. The keywords are searched from top to bottom in the learner preference tree. If the keyword is not found it adds as a leaf node and calculate the rating of the new node’s all predecessor.

4. Rating Prediction

Two learners’ with similar attribute keywords are considered as neighbors and calculate the similarity between two using cosine similarities method to overcome the sparsity problem. The user’s high rating resources are recommended to
the learners.

B. Implicit Attribute-Based Module

The time complexity of executing the collaborative filtering algorithm grows linearly with the number of items and the number of users. The recommendation algorithm will suffer serious scalability problem when the number of users and resources increase tremendously. In general, Genetic algorithm are believed to be effective on NP-complete global optimization and it can provide good near-optimal solutions in the reasonable time. Therefore, this research uses a GA for optimization of implicit attributes’ weight. In the attribute space, different people may place different emphases on interrelated attributes. The goal of a GA is to find the relationship between the overall rating and the underlying attributes’ rating for each learner. More specifically, given the rating data of a learner, a GA computes his/her preference model in terms of the implicit attributes’ weight. Truly, GA used as a supervised learning task whose fitness function is the mean absolute error (MAE) of the RS.

1. Genetic Algorithm Method

A genetic algorithm is a search technique used in computing to find exact or approximate solutions to optimization and search problems. It is after the model of the nature ability to adapt environment. The process of genetic algorithm as shown in fig.5.3 is given as follows

Step 1 Generate initial population.
Step 2 Fitness function used to evaluate each individual.
Step 3 Iterate until stopping criteria
Step 3.1 Selections and Crossover
Step 3.2 Mutation
Step 3.3 Evaluate each individual using fitness function.
Step 3.4 Make next generation
Step 3.5 Evaluate stopping criteria.

![Flowchart for Genetic Algorithm](image)

2. Implicit Attribute Weight Extraction

The goal of a GA is to find the relationship between the overall rating and the underlying attributes’ rating for each learner. Genetic Algorithm computes his/her preference model in terms of the implicit attributes’ weight.

3. Coding Strategy

As shown in Fig. 5.4 the chromosome scheme represents the implicit attributes’ weights for the users and the items where \( U_i = (w_{i1}, w_{i2}, ..., w_{ik}) \) & \( I_j = (e_{j1}, e_{j2}, ..., e_{jk}) \) indicates the implicit attributes’ weight vector for user i and item j respectively and k with the number of attributes\( \sum_{k=1}^{K} e_{j1} = 1, \sum_{k=1}^{K} w_{ij} = 1 \) indicates the implicit attributes’ weight vector for user i and item j respectively and k with the number of attributes. Each chromosome has N + M rows corresponding with N users and M items, and also has K*10 columns corresponding with K implicit attributes. Each attribute weight is coded in the form of a binary string. It is represented as 0 to 1 continuously.

![Implicit attributes’ weight vectors chromosome coding strategy](image)

4. Fitness Function

In genetic algorithm, a fitness function is used to evaluate the superiority of the each individual and the higher score indicates that the individual is better. It use a Genetic Algorithm as a supervised learning task whose fitness function is the mean absolute error (MAE) of the Recommender system. The MAE is obtained by comparing the real ratings with the Predicted ratings made based on two matrices. The fitness function can be calculated by using the following formula, in this the real rating of the item j given by the user is represented as \( r_{ij} \) and the weight of the attribute k for user i and j are represented as \( w_{ik}, e_{jk} \) and \( M_i \) and number of rated items are represented as:

\[
\text{Fitness} = \sum_{i=1}^{N} \left( \sum_{j=1}^{M_i} \left| \sum_{k=1}^{K} w_{ik} e_{jk} R_{ij} - r_{ij} \right| \right)
\]

5. Selection

A probabilistic selection is performed based on the individual’s fitness such that the better individuals have an increased chance of being selected. The sum of the fitness in a population is constant, an individual with lower fitness (higher prediction accuracy) has a larger probability to be chosen. The universal sampling method scheme yields a good individual to be selected for reproduction of the next population. The selection probability can be calculated by using the following formula, in this the value of fitness function for chromosome c is represented as fitness \( S_c \) and PS represents the number of individual in the population and \( P_e \) represents the selection probability for chromosome c.

\[
P_e = 1 - \frac{\text{fitness } S_c}{\sum_{c=1}^{PS} \text{fitness } c}
\]

6. Crossover and Mutation
The one-point crossover technique is used to produce the offspring. Based on this technique strings are selected randomly on both parents’. This operator is implemented for each row of the chromosome that presented in Fig 5.4 to separate each attribute. The Single point mutation technique is used to introduce the diversity of recommendation. Mutation operator is used to investigate some of the unvisited points in the search space, and also avoid the premature convergence of the entire feasible space caused by some super chromosomes.

7. Rating Prediction
After the implicit attributes’ weight optimization, the similarity between learners using the Implicit Attribute Based method can be calculated by using the cosine similarity.

8. Recommendation
The top N learning resources with the higher predicted rate are considered as the recommendation results. To improve the quality of recommendations create a hybrid of two methods by the weighted combination method. A linear combination of Explicit Attribute Based collaborative filtering and Implicit Attribute Based collaborative filtering is used for recommendation (EB-IB-CF).

VI. EVALUATION METRICS
In this paper the recommendation algorithms are evaluated by using the Decision support accuracy metrics. It assumes the prediction process as a binary operation which represents the items predicted is good or bad. The most popular metrics in this category are precision and recall. It can be defined as follows which is used to find the accuracy of the recommendation.

Precision = \( \frac{tp}{tp+fp'} \)
Recall = \( \frac{tp}{tp+fn'} \)

Where tp represents the true positive and fp represents the false positive and fn represents the false negative.

VII. RESULT AND DISCUSSION
In this recommendation method based on the number of similar neighbors and recommendations the performance and quality of the process increases when compared to the traditional approaches. As shown in graph 7.1 Except for the content based algorithm, the number of similar neighbor increases the F1 of each algorithm also increases. When this number is increases to a certain point the precision of each algorithm begins to decreases because at a certain point the several dissimilar users are considered as similar users in the collaborative based algorithm due to this the accuracy of recommendation will decrease. But in our approach the threshold value set for similar users which increase the quality of the system.

As shown in graph 7.2 based on the recommendation resources increases the F1 of each algorithm decreases. But in this approach it provides the better performance when the recommendation resources is small when compared to the other algorithms because it provides more advantages by combining the three kinds of information namely multidimensional attributes of resources, user’s ratings and implicit attributes.

VIII. CONCLUSION AND FUTURE WORK
Personalization and recommendation of learning resources is one of the most important applications of recommendation systems in an e learning environment. This paper provides novel personalized recommendation framework that utilizes implicit and explicit attributes of resources which eliminate the sparsity and cold start problems. It improves the quality of recommendation by providing diversity of recommendation list and provides suitable resources to the learners.
REFERENCES


