A Web Based Recommendation System for Personal Learning Environments Using Hybrid Collaborative Filtering Approach

Dr.K.Anandakumar[1], K.Rathipriya[2], Dr.A.Bharathi[3]

Associate Professor (SL.G), Department of Computer Applications, Bannariamman Institute of Technology, Sathyamangalam, India[1].

Assistant Professor, Department of Information Technology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India[2].

Professor, Department of Information Technology, Bannariamman Institute of Technology, Sathyamangalam, India[3].

ABSTRACT: The large growth of Web has influenced the generation of huge e-learning resources. This work is focused to devise a personal recommendation system that will address the sparsity and cold-start problems and that will provide a have a more diverse recommendation list for each learner. Here Improved Neighborhood- based Collaborative filtering and Hybrid Genetic algorithm with Particle Swarm Optimization (PSO) method is implemented. These techniques are employed for improving the diversity, and the convergence towards the preferred solution taking into account the preferences of users. The results obtained from the experiments show that the proposed method outperforms current algorithms in terms of accuracy measures and can alleviate cold-start and sparsity problems and generate a more diverse recommendation list as well.

KEYWORDS: Personal learning Environments, Collaborative Filtering, Particle Swarm Optimization, Genetic Algorithm, learning behaviour.

I. INTRODUCTION

In the education field , the learning behaviour of a learner is a criteria of greater importance since it greatly affects the learning process[1].Also, the process of learning is based on the teaching behaviour as well; if the teacher uses the same set of methods for teaching then the learner will not succeed [2]. Several learning processes are available; one of the learning process is based on mobile learning. Mobile learning process marked by flexibility, customization and collaboration [3 and 4].

One of the most important method is personalized recommendation and it is prosed by [5, 6], these methods implement the knowledge discovery methods for learners such as data mining and machine learning. These methods are used to discover the interests of users based on their activities and these activities are used to make the recommendation [7, 8].

Some of the example for e-learning sources for personalized recommendation is book recommendation in Amazon.com [9], movie recommendation in Netflix.com [10] and video recommendation in Tivo.com [11].Nowadays, there is a rapid development in web based education system. With the rapid development of online learning system, large numbers of e-learning sources are generated [12].

Some of the general procedures for personalized recommendation system consist of three steps; these steps are used to solve the problem in this system and are proposed in [13]. These steps are (i) record the user behaviour and generate a template for the behaviour (ii) Behaviour template should be maintained (iii) Generate recommendation information to users based on their behaviour template.

A Recommender system is to provide a personalized recommendation on the utility of a set of objects belonging to a given domain, starting from the information available about users and objects [22].

The given terms are used to define the recommendation problem$C \times I \rightarrow R$, in this expression C denotes the set of users of the system. I denote the recommender system and R denotes the totally ordered system [23]

The paper is organized in this way: Section 2 introduces some background study about recommender systems; Section 3 describes the proposed methodology; Section 4 gives the result of the recommender system. Section 5 gives the conclusion of the research work.

II. RELATED WORK

A Rule based recommendation system creates user behaviour rules through user history by association rule mining technology [14]. It is one of the simple recommendation system but it is hard to control the quality and quantity of the rules [15]. A Content-based recommendation is generated depending on the similarity differences between the content of resources and the preferences of the users [16].

Another approach for recommendation system is collaborative filtering studied in [17] and it is one of the most dominant methods. This system is not dependent on the content of the resources and the user's new interests. The two types of collaborative filtering are memory based and model based algorithm [18].

There are three ways to find out the student interaction and assessments to improve the learning [19]. The three steps are monitoring tools, metacognitive tools and guide systems. Monitoring tools are used to show or monitor the behaviour of the student to the student and the teacher based on their approach [20]. Metacognitive tools analyse the student behaviour based on desired outcome [21].
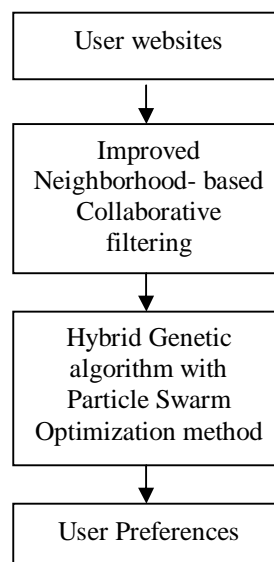
Fig. 1 Flow diagram of the Proposed methodology

The above figure (Fig. 1) represents the steps of the proposed methodology. Initially log data from User websites are given as input. Then Improved Neighbourhood based Collaborative Filtering Algorithm is used to classify similar users. After which Hybrid GA with PSO is used optimize the search process after which user preferences indicating final recommendations are generated.

III. PROPOSED METHODOLOGY

A. Improved Neighborhood-based Collaborative Filtering

The most important approach to Collaborative filtering is the neighborhood-based approach. It is very popular and easy to implement. But however it has some drawbacks that are similarity function and the problem in previous neighborhood-based methods is that they do not account for interactions among neighbors. To overcome this limitation, improved Neighborhood-based collaborative filtering is used, the aim of which is to improvise the previous methods in order to remove the global effects.

This strategy is to estimate one "effect" at a time, in sequence. At each step, we use residuals from the previous step as the dependent variable for the current step. Consequently, after the first step, the r_{ui} refer to residuals, rather than raw ratings.

For each of the effects mentioned above, our goal is to estimate either one parameter for each item or one parameter for each user. For the rest of this subsection, we describe our methods for estimating user specific parameters; the method for items is perfectly analogous.

Denoted by x_{ui} the explanatory variable of interest corresponding to user u and item i. For user main effects, the $x_{ui}'s$ are identically 1. For other global effects, we center x_{ui} for each user by subtracting the mean of x_{ui} for that user. In each case, model is described as:

$$r_{ui} = \theta_u x_{ui} + error$$

With sufficient ratings for user u, we might use the unbiased estimator:

$$\hat{\theta}_u = \frac{\sum_i r_{ui} x_{ui}}{\sum_i x_{ui}^2}$$

Where each summation is over, all items rated by u. However, for sparse data, some values of $\hat{\theta}_u$ may be based on very few observations, thereby resulting in unreliable estimates.

To avoid overfitting, we shrink individual values of $\hat{\theta}_u$ towards a common value. Shrinkage can be motivated from a Bayesian perspective. Suppose that the true $\hat{\theta}_u$ are independent random variables drawn from a normal distribution

$$\hat{\theta}_u \sim N(\mu, \tau^2)$$

For known μ and τ^2 while

$$\hat{\theta}_u | \theta_u \sim N(\theta_u, \sigma_u^2)$$

For known σ_u^2. Then, the best estimator for θ_u is its posterior mean:

$$E(\theta_u | \hat{\theta}_u) = \frac{\tau^2 \hat{\theta}_u + \sigma_u^2 \mu}{\tau^2 + \sigma_u^2}$$

a linear combination of the empirical estimator $\hat{\theta}_u$ and the common mean μ. The parameter σ_u^2 can be estimated from the formula for the variance of a weighted mean, while μ can be estimated by the mean of the $\hat{\theta}_u$'s (optionally weighted by nu). Empirical Bayes suggests that the maximum likelihood estimate of τ^2 can be found as the solution to:

$$\tau^2 = \frac{\sum_u \left[(\hat{\theta}_u - \mu)^2 - \sigma_u^2 \right] / (\tau^2 + \sigma_u^2)^2}{\sum_u (\tau^2 + \sigma_u^2)^{-2}}$$

Slightly simpler estimator for θ_u by assuming μ = 0 and σ_u^2
is proportional to $1/n_u$, which yields

$$\frac{n_u \hat{\theta}_u}{n_u + \alpha}$$

In the above equation n_u is the number of ratings by user u and α is a constant.

B. Hybrid GA with PSO

Many machine learning approaches have been applied by learning module developers, including inductive logic programming, decision tree, deductive reasoning, fuzzy logic, neural networks, and Genetic Algorithm.

The optimization problem is a difficult task to be carried out due to the huge search space. To overcome this setback, genetic algorithm is introduced to obtain the global optimum of a function. Genetic algorithm performs the optimization in four stages that are initialization, selection, crossover, and mutation [24].

The genetic algorithm starts with the initial solutions that are selected from a set of configurations in the search space called population using randomly generated solutions. By using a user fitness function, an intial solution of the each user can be examined. The second stage is the selection stage that which has the high score in the fitness is to be selected and reproduced by itself. The third stage is the crossover stage to produce the two chromosomes called as children produced from the couple of chromosomes called as parents. Crossover stage operates by swapping corresponding segments of a string representation of these parent and children. The final stage is the mutation stage which operates on a single chromosome.

Pseudo code of the proposed hybrid algorithm

Step 1: Initialization

$$g \leftarrow 0$$
$$\text{for } i - 1 \text{ to M do}$$

Initialize individuals x_i to random values

$$F_1 \leftarrow f(x_i)$$
$$\text{end for}$$
$$Pop \leftarrow \{x_1, x_2, \ldots, x_M\}$$
$$F \leftarrow \{F_1, F_2, \ldots, F_M\}$$

{Main Loop}
While (not terminate condition)
    {Genetic Operators}
   Pop: Selection process
    Initialize parameters;
Initialize swarm;
Evaluate the Performance of particles;
Do {
Find the personal best;
Find the global best;
Update the velocity;
Update the position;
Evaluate the Performance of particles;
} While (Termination criterion not satisfied);
   $\textbf{\textit{Pop}} \leftarrow \textbf{\textit{Crossover}}(\textbf{\textit{Pop}}, \textbf{\textit{p}}_c)$
   $\textbf{\textit{Pop}} \leftarrow \textbf{\textit{Mutation}}(\textbf{\textit{Pop}}, \textbf{\textit{p}}_m)$
{Evaluation Loop}
  for i=1 to M do
    $\textbf{\textit{F}}_i \leftarrow \textbf{\textit{f}}(\textbf{\textit{x}}_i)$
  end for
$$\textbf{\textit{F}} \leftarrow \{\textbf{\textit{F}}_1, \textbf{\textit{F}}_2, \ldots, \textbf{\textit{F}}_M\}$$
$$\textbf{\textit{g}} \leftarrow \textbf{\textit{g}} + \textbf{1}$$
end while

Particle Swarm Optimization (PSO) has a wide range of applications in many areas and it is one of the population based stochastic optimization techniques [25]. PSO is used to search the particles for optimal solution. Foe every iteration, update the two "best" values of each particle. pbest is the first best value; it represents the position vector of the best solution and it is achieved by this particle. The second best value is gbest which is the global best position and

it identifies the best position found by the entire swarm until the iteration t. The particles update the velocity and position after finding these two best values and it is given by the equation below

$$v_{id}^{\vec{t}+1} = wv_{id}^{\vec{t}} + c_1 \times rand_1 \times \left(pbest_{id}^{\vec{t}} - \vec{p}_{id}\right) + c_2 \times rand_2 \times \left(gbest_{gd}^{\vec{t}} - \vec{p}_{id}\right)$$

$$p_{id}^{\vec{t}+1} = p_{id}^{\vec{t}} + v_{id}^{\vec{t}}$$

In the above two equations, the velocity of the id particle is denoted by $v_{id}^{\vec{t}}$ at t[th] iteration, current solution of the id[th] particle is given by $p_{id}^{\vec{t}}$. $rand_1 \ and \ rand_2$ are random numbers generated uniformly between 0 and 1. $c_1 \ and \ c_2$ denotes the self-confidence and swarm confidence factor

IV. EXPERIMENTAL RESULTS

There are various types of datasets which can be recommended for this system. Here user's opinion is collected from various users by creating a synthetic dataset and is taken for the process where it is implemented in java platform.

The performance of this approach is measured using Decision support accuracy metrics assume the prediction process as a binary operation; either items are predicted (good) or not (bad). Precision and recall are the most popular metrics in this category. When referring to recommender systems, recall and precision can be defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

Where, TP – True Positive, FP – False Positive and FN- False Negative. From the precision and recall the f-measure is calculated.

$$\text{FMeasure} = 2.\frac{\text{Precision.Recall}}{\text{Precison} + \text{Recall}},$$

Then the Mean Absolute Error (MAE) can be calculated by

$$\text{MAE} = \frac{\sum_{i=1}^{i=S} | \text{R}^i - \text{P}^i|}{|\text{S}|}$$

where P^i is the predicted rating for resource i, R^i is the learner given rating for resource i, and S is the total number of the pair ratings R^i and R^i.

To make a prediction for a target user, here considered the number of rated items by the target user in training sets 5, 10, and 20. Then, we calculated the corresponding MAE using different algorithms.

V. TABLE 1 PERFORMANCE EVALUATION UNDER MAE

Techniques	MAE1	MAE5	MAE10
Content based	2.59	2.32	2.18
Existing[26]	2.53	2.19	2.09
Proposed	2.49	2.15	2.05

The above table(TABLE 1) shows the performance comparison of the proposed system with the previous techniques .The performance evaluation is done by measuring the MAE.A decrease in the MAE value indicates an improved

prediction accuracy of the recommendations generated. Here the proposed technique is compared with existing and content based system.

VI. CONCLUSION

Personalization is a factor which is to be given much importance as far as E-learning Recommender systems are concerned. The personalized recommendation system for e-learning applications is developed here by implementing Improved Neighborhood- based Collaborative filtering and Hybrid Genetic algorithm with Particle swarm optimization. The proposed system is evaluated using MAE and F- measure. The existing and proposed methods are compared and the results indicate that the proposed technique generates recommendation results with improved accuracy by minimizing the error rate.

REFERENCES

1. Mangione, G. R., Gaeta, M., Orciuoli, F., & Salerno, S.,"A Semantic Metacognitive Learning Environment", Cognitive and Metacognitive Educational Systems, AAAI Fall Symposium, 2010.
2. Looi, C.-K., Wong, L.-H., So, H.-J., Seow, P., Toh, Y., Chen, W., et al.,"Anatomy of a mobilized lesson: learning my way", Computers & Education,Vol. 53, No.4, pp. 1120–1132,2009.
3. Cobcroft, R., Towers, S., Smith, J., & Bruns, A.," Mobile learning in review: Opportunities and challenges for learners, teachers, and institutions". Proceedings of Online Learning and Teaching , pp.21-30, 2006
4. Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M.,"Literature review in mobile technologies and learning". NESTA Futurelab report 11,2004.
5. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.," GroupLens: an open architecture for collaborative filtering of netnews", ACM Conference on Computer Supported Cooperative Work , pp. 175–186. 1994
6. W. Hill, L. Stead, M. Rosenstein, G. Furnas., "Recommending and evaluating choices in a virtual community of use", Human Factors in Computing Systems, pp. 194–201, 1995.
7. Resnick,P., Murray Hill ,N.J., Varian,H.R., " Recommender systems", Communications of the ACM,Vol.40,No. 3., pp. 56-58, 1997.
8. Liu,J.,Shi,K., Guo,Q., "Solving the accuracy–diversity dilemma via directed random walks", PhysRevE.85.016118,2012.pp. 1-9,2012
9. Linden,G., Smith,B., York,J, "Amazon.com recommendations: item-to-item, collaborative filtering", IEEE Conference on Internet Computing,Vol. 7,No.1.,pp.76–80,2003.
10. Bennett, J., Lanning, S., "The Netflix prize", in: 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007.
11. Ali, K.,W. van Stam, TiVo: "Making show recommendations using a distributed collaborative filtering architecture", Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 394–401, 2004.
12. Chen,W., Niu,Z., Zhao,X and Li,Y., "A Hybrid Recommendation Algorithm Adapted in E-Learning Environments," World Wide Web,Vol.17,No.2.,pp.271-284, 2012,
13. Ricci, F., Rokach, L., & Shapira, B.," Introduction to Recommender Systems Handbook, Recommender Systems Handbook", Springer, pp.1–35,2011.
14. Smyth, B., McCarthy, K., Reilly, J., O'Sullivan, D., McGinty, L., & Wilson, D.," Case studies in association rule mining for recommender systems".International Conference on Artificial Intelligence,pp.1-7,2005.
15. Agrawal, R., Imielinski, T., & Swami, A." Mining association rules between sets of items in large databases". In Proceedings of ACM International Conference on Management of data, Vol.22, No.1, pp. 207–216, 1993.
16. Symeonidis, P." Content-based dimensionality reduction for recommender systems". In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), Data analysis, machine learning and applications, studies in classification, data analysis, and knowledge organization. pp. 619–626, 2008.
17. Lu, Y. T., Yu, S. I., Chang, T. C., & Hsu, J. "A content-based method to enhance tag recommendation". Twenty-first International joint conference on artificial intelligence, pp.2064-2069, 2009.
18. Tatu, M., Srikanth, M., & Silva, T. D. "Tag recommendations using bookmark content". European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", pp.96-107,2008.
19. Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. "From mirroring to guiding: A review of state of the art technology for supporting collaborative learning". International Journal of Artificial Intelligence in Education, Vol. 15, No.4. pp.261–290,2005.
20. Kahrimanis, G., Meier, A., Chounta, I.-A., Voyiatzaki, E., Spada, H., Rummel, N., et al. "Assessing collaboration quality in synchronous CSCL problem-solving activities: Adaptation and empirical evaluation of a rating scheme .Learning in the synergy of multiple disciplines". 4th European conference on Technology Enhanced Learning, pp. 267–272, 2009.
21. Gaudioso, E., Montero, M., Talavera, L., & del Olmo, F. H. "Supporting teachers in collaborative student modeling: A framework and an implementation". Expert Systems with Applications, Vol. 36, No.2, pp. 2260–2265, 2014.
22. Nicola Capuano, Matteo Gaeta, Pierluigi Ritrovato and Saverio Salerno. "Elicitation of latent learning needs through learning goals recommendation", Computers in Human Behavior, Vol. 30 , pp.663–673,2014.
23. Adomavicius, G., & Tuzhilin, A. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, pp. 734–749, 2005.
24. Davis, L. " Handbook of genetic algorithms".Vol.3,No. 4,pp. 446-448,1991.

25. Kennedy, J., Eberhart, R.C." Particle swarm optimization", IEEE International Joint Conference on Neural Networks,Vol.4, pp. 1942–1948,1995

BIOGRAPHY

Dr.K.Anandakumar is an Associate Professor in the Computer Applications Department, Bannariamman Institute of Technology, Sathyamangalam, Erode District, TamilNadu, India. He received Master of Computer Application (MCA)degree in 1998 and Ph.D degree in 2012 from Bharathiar University,TamilNadu, India. His interests include Data mining and warehousing, Imageprocessing,NeuralNetworks,etc.

K.Rathipriya is an Assistant Professor in the Information Technology Department, Avinashilingam University forWomen, Coimbatore, and TamilNadu. She received Master of Information Technology (MSc) degree in 2009 from Bharathiar University, TamilNadu. She is pursuing her Master of Philosophy (M.Phil) degree in Computer Science in Computer Science Department, Dr.S.N.S Rajalakshmi College of Arts & Science, Coimbatore, TamilNadu, India. Her interests include Data Mining and warehousing, Recommender systems, etc.

Dr.A.Bharathi is a Professor in the Information Technology Department, Bannariamman Institute of Technology, Sathyamangalam, Erode District, TamilNadu, India. She received Master of Engineering (M.E) degree in Computer Science & Engineering from Bannari Amman Institute of Technology in 2007 and Ph.D degree in Information and Communication Engineering from Anna University, Chennai in 2012. Her interests include Software Engineering, Data Mining and Warehousing, Neural Networks, Image processing etc.