Available online at www.sciencedirect.com

SciVerse ScienceDirect

ELSEVIER

IERI Procedia 2 (2012) 571 – 576

Procedia IERI

www.elsevier.com/locate/procedia

2012 International Conference on Future Computer Supported Education

A New Recommendation Approach Based on Implicit Attributes of Learning Material

Mojtaba Salehi[a]*, Isa Nakhai Kmalabadi[a], Mohammad Bagher Ghaznavi Ghoushchi[b]

[a]*Faculty of Engineering, Tarbiat Modares University, 14115, Tehran, Iran*

[b]*Department of Electrical Engineering, Shahed University, Tehran, Iran*

Abstract

A personalized recommendation is an enabling mechanism to overcome information overload occurred in the new learning environments and deliver suitable learner materials to learners. But recommender system technology suffers from some problems such as cold-start and sparsity. Since users express their opinions based on some specific attributes of materials, this paper proposes a new recommender system for learning materials based on their attributes to address these problems. Weight of implicit or latent attributes for learners is considered as chromosomes in genetic algorithm then this algorithm optimizes the weight of implicit attributes for each learner according to historical rating. Then, recommendation is generated using Nearest Neighborhood Algorithm (NNA). The experimental results show that our proposed method outperforms current algorithms and can perform superiorly and alleviates problems such as cold-start and sparsity.

Keywords: collaborative filtering, e-learning, sparsity, Personalized Recommendation Learning Material, E-learning

1. Introduction

With the growth of technology in educational organizations at recent years, Web-based learning environments are becoming very popular. One of the important parts in the new learning environments is recommender system. A recommender system in an e-learning context is a software agent that tries to ”intelligently” recommend actions to a learner based on the actions of previous learners. One of the most

* Corresponding author. Tel.: +98-2182884323; fax: +98-2182884323, *E-mail address:* Mojtaba.salehi@modares.ac.ir.

2212-6678 © 2012 Published by Elsevier B.V. Selection and peer review under responsibility of Information Engineering Research Institute
Open access under CC BY-NC-ND license. doi:10.1016/j.ieri.2012.06.136

important applications of recommender systems in learning environments is materials recommendation. RSs use opinions of a community of users to help individuals identify material of interest from a potentially overwhelming set of choices more effectively. By using material RSs in learning environments, we can address two problems, personalization and information overload. In this situation, recommender system offers which learning objects should learners study next (Klasnja-Milicevic et al., 2011).

Recommender systems that have been deployed usually in e-commerce entities for expressing customer's interests use three strategies for recommendation including content-based (Khribi et al., 2009), collaborative filtering (Liu and Shih, 2007; Romero et al., 2009; Luo, et al., 2010), and hybrid recommendation (García, et al., 2011). While the recommender system algorithms try to address information overload and personalization problem, with growing numbers of existing users and materials tremendously, these algorithms will suffer serious scalability and sparsity problems. In addition, most of these algorithms don't consider attributes of materials that can address these problems. We can consider two groups of attributes, explicit attributes and implicit (latent) attributes, for learner and learning material. Explicit attributes are known such as subject, education type for learning material, but implicit attributes that are latent can be can be inferred by historical rating. Some researches tried to use these two types of attribute for recommendation simultaneously by combining attributes (feature) of users or materials with historical rating. Robin (2002) reviewed several hybrid recommender methods developed to combine the external features and historical rating data for higher predication accuracy. According to the experiment results reported, it is believed that both features and the historical ratings have great values to estimate the predication function for recommendation.

In order to generate recommendations with higher accuracy and address some problems in existing recommender system algorithms such as sparsity, this research uses implicit attributes of user and item for recommendation. Genetic algorithm is used for extracting implicit attributes of leaner from historical rating in the shape of weight vectors. Then, NNA produces recommendations by the produced attributes weight vectors. The main contribution of this paper is improving the quality of recommendation and addressing sparsity problem using explicit attributes. Rest of this paper is organized as follows: Methodology Section introduces describes the proposed mechanism step by step. Experiment section applies the proposed algorithm for a datasets to evaluate the performance. Finally, Conclusion section provides the concluding remarks.

2. Methodology

In the proposed mechanism is presented in this section step by step.

2.1. Implicit attributes

Learning materials usually have several kinds of attributes. Therefore, in order to consider learner's preference accurately, attributes of learning materials should be taken into account as much as possible. Since the ratings depend on the attributes of learners and learning materials, the rating function could be denoted as $\varphi(M,\vec{U},\vec{I})$, M is a prediction model learned from the historical rating data H. $\vec{U}$ and $\vec{I}$ are attributes of the learners and learning materials, respectively. Unfortunately, in most cases we can't use the mentioned model. Because the selection of suitable attributes for the learner and material in a CF problem is an almost impossible mission, which needs a lot of prior expert knowledge in the art fields rather than technology in most cases (Zhong and Li, 2010). However, we can use the historical rating data in a user-material matrix for discovering some of valuable attributes of learner and learning material that are called implicit attributes reflected characteristics of learning material and learner. Thus, we can use the predication models built based on the observed attributes and latent or implicit attributes to improve recommendation process for acquiring the higher prediction accuracy (Zhong and Li, 2010). In this section, genetic algorithm is used to find the relationship between the overall rating and the underlying implicit attribute weight vector for each learner.

2.2. Implicit Attribute optimization

The GA mimics the process of natural evolution by combining the survival of the fittest among solution structures with a structured, yet randomized, information exchange and creates offspring (Salehi and Tavakkoli-Moghaddam, 2010). In the following of this section, we describe GA process step by step:
Coding: In this study, each chromosome that has been shown in the Fig. 1 represents the implicit attributes weights for users and materials where $w_i = (w_{i1}, w_{i2}, ..., w_{iK})$ and $e_i = (e_{i1}, e_{i2}, ..., e_{iK})$ indicate attributes weight vector for user *i* and material *i* that *K* is number of defined implicit attributes and $\sum_{j=1}^{K} w_{ij} = 1$, $\sum_{j=1}^{K} e_{ij} = 1$.

	1										…….	*K*									
U(i) or I(i)	1	1	1	0	0	0	0	1	0	0	……..	0	1	0	0	0	1	1	1	1	0

Fig. 1. Representation of attributes weight vectors in a chromosome

Two matrixes of attributes weight $W_U = (w_1, w_1, ..., w_N)^T$ and $W_I = (e_1, e_1, ..., e_M)^T$ that indicate attributes weight vectors for *N* users and *M* materials respectively become the optimizing targets. Its initial solution could be some random values gained by an off-line process.
Fitness function: When individual is applied to generate recommendation's results, the similarity between final forecasting results with the actual rating values can express its forecasting accuracy. It's the basis of fitness. So, the accuracy function is defined as follows.

$$f(W_U, W_I) = \sum_{i=1}^{N} \mathrm{f}(\mathrm{w_i}) = \sum_{i=N}^{N} \sum_{\mathrm{j=1}}^{\mathrm{M_i}} \left| \sum_{k=1}^{K} w_{ik} . e_{jk} - r_{ij} \right| \qquad (1)$$

Where r_{ij} is actual rating of material *j* by user *i*, w_{ik} and e_{jk} are weight of attribute k for user and material respectively and M_i is number of rated materials by user *i*. When $f(W_U, W_I)$ is lower, the forecasting accuracy would be higher.
Selection operation: A probabilistic selection is performed based upon the individual's fitness such that the better individuals have an increased chance of being selected. Here, the universal sampling method is adopted for selecting the good strings and the probability of selecting each string is calculated by:

$$p_c(W_U, W_I) = 1 - \frac{f_c(W_U, W_I)}{\sum_{C=1}^{PS} f_C(W_U, W_I)} \qquad (2)$$

Where $f_c(W_U, W_I)$ denotes the value of fitness function for chromosome c, PS is number of individuals in the population or population size and $p_c(W_U, W_I)$ denotes the probability of selecting chromosome c.
Crossover and mutation operation: In this work, one-point crossover is used to produce offspring. Single crossover point on both parents' strings is selected randomly. Mutation operator makes random changes in one or more elements of the string. According to mutation rate, randomly selecting some elements of individual, and changing its value, new individual can be gained.

2.3. Rating prediction

After implicit attribute weight optimization, similarity degree between learners by implicit attribute based (IAB) can be calculated by following formula that is cosine similarity:

$$sim_{IAB}(U_a,U_b)=\frac{\sum_{i=1}^{K} w_{ai}.w_{bi}}{\sqrt{\sum_{i=1}^{K} w_{ai}^2}.\sqrt{\sum_{i=1}^{K} w_{bi}^2}} \tag{3}$$

The predication rating of learning material i by U_a using implicit attribute based method is $P_{IAB}(U_a,i)$ that is gained by the rating of U_a neighbourhood, $N_{IAB}(U_a)$, that have rated i before. The computation formula is as the follows:

$$P_{IAB}(U_a,i)=\overline{R}_{U_a}+\frac{\sum_{j\in N_{IAB}(U_a)} sim_{IAB}(U_a,U_j)\times(R_{U_j}(i)-\overline{R}_{U_j})}{\sum_{j\in N_{IAB}(U_a)} sim_{IAB}(U_a,U_j)} \tag{4}$$

Where $\overline{R}_{U_a}$ and $\overline{R}_{U_j}$ average rating of learning materials rated by active learner U_a and U_j respectively and $sim_{IAB}(U_a,U_j)$ is the similarity between active learner U_a and U_j that is a member of $N_{IAB}(U_a)$.

3. Experiments

The effectiveness of proposed approach in terms of recommendation quality is investigated as follows:

3.1 Evaluation metrics and Data set

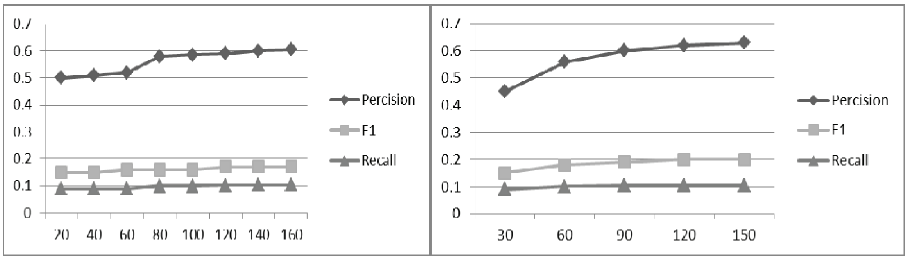
In order to check the performance of the proposed algorithm, a real-world dataset is applied in our simulations. MACE† dataset that is pan-European initiative to interconnect and disseminate digital information about architecture is used for experiment. This dataset is issued from MACE project that is done from September 2006 to September 2010. This dataset contains 1148 learners and 12000 materials.
Recall and precision have been used by various researchers to evaluate recommender systems (Pazzani and Billsus, 2007, Herlocker, 2004). In this research we use this measure for evaluation. Since increasing the size of the recommendation set leads to an increase in recall but at the same time a decrease in precision, we can use F1 measure (Shih and Liu, 2008).

3.2 Parameters setting

The optimal values of crossover and mutation probabilities are problem specific that often is obtained by trial and error. Therefore According to our experiments on the impacts of various combinations of pc and pm, we set $pc=0.85$ and $pm=0.17$.
Fig. 2 shows the performance results for optimal values of $pc=0.85$ and $pm=0.17$ for implicit attribute-based recommendation by varying the number of generations and population size. Fig. 2 indicates that the performance is gradually improved generation by generation. However, the improvement becomes insignificant when the number of generations is greater than 70. As expected, the performance improves as the population size increases, but it reaches a saturation point at $PS=0.90$ and any further increment does not improve results very much. Therefore, we can fix the population size to 90.

---

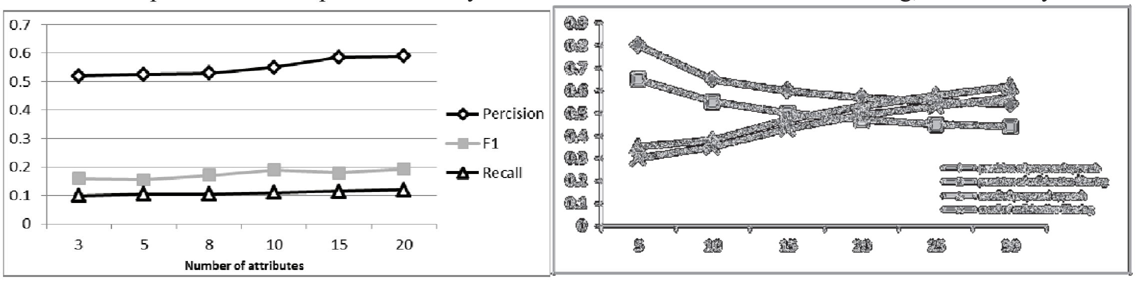
† - Metadata for Architectural Contents in Europe



(a) (b)

Fig. 2. Performance results of implicit attribute- based recommendation vs. a) generation number, b) population size

One of the other parameters for the implicit attribute-based recommendation is number of attributes, *K*. Fig. 3(a) shows the results obtained for the proposed model with different number of attributes, where the minimum number of rating required for test learners, *M*, was 50 and the number of learner, *N*, was 500. It can be seen that the performance improves steadily with the number of attributes increasing, but not very much.



(a) (b)

Fig. 3. (a) Performance of the method with respect of K, b) Comparison of the method and CF based on number of recommendation

3.3 Performance comparison

In experiments, each dataset is ordered by learners' access timestamp, and then is divided into a training set and a test set. The algorithm is then trained on the training set and top N- learning materials are predicted from that learners' test set. To evaluate the sensitivity of different recommendation numbers, we compare proposed approach and collaborative filtering based on number of recommendation for recall and precision measure that is presented in Fig. 3(b), N=500, K=8, and M=50. As expected, when the number of recommendations increases, the precision drops smoothly but the recall improves gradually. The results demonstrate the effectiveness of the proposed approach.

Table. 1. A comparison of prediction accuracy of various methods

Method	Error		
	MAE	RMS	0/1 loss
Proposed method(Attribute based)	0.807	1.093	63.2
User based	0.873	1.172	64.7
Mixture pLSA	0.848	1.170	63.4
CR [31]	0.994	-	-
BC [31]	1.103	-	-
BN [31]	1.066	-	-

Table 1 presents the experimental results obtained by the proposed method, the memory-based method, Gaussian pLSA mixture method Hofmann (2004) and results published in Breese et al. (1998). Since the data set will influence the results of CF algorithm, comparing of different algorithms is difficult. For the mixture pLSA, results are chosen the best results in Hofmann (2004). The results of learner-based and proposed method obtained from the same data set. Comparisons were produced for N=500 learners with the average number of ratings about 100, and M= 50. As can be seen, the proposed multi-attribute based method has better prediction accuracy of the memory-based, mixture pLSA method and other methods in terms of MAE.

4. Conclusions

This paper presents a novel personalized recommender system that utilizes implicit attributes of materials to address the drawback of existing recommendation models. In this work, GAs is used for attribute weight optimization to solve sparsity problem. The experiment results show that the proposed approach performs better than traditional collaborative filtering. The main contribution are improving the quality of recommendations and addressing sparsity problem by considering of implicit attributes for recommendation.

References

[1] Breese, J.S., Heckerman, D., Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the 14th conference on uncertainty in artificial intelligence, 1998:43–52.

[2] Claypool, M., Gokhale, A., Combining content-based and collaborative filters in an online newspaper. In SIGIR ’99 workshop on recommender systems: Algorithms and evaluation. 1999;195–204.

[3]García, E., Romero, C., Ventura, S., Castroa, C.d., A collaborative educational association rule mining tool, The Internet and Higher Education, 2011;14(2):77-88.

[4] Hafmann, T. Latent semantic models for collaborative filtering. ACM Transactions on Information Systems, 2004;22: 89–115,

[5] Herlocker, J.L. et al., Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 2004;22(1):50–53.

[6] Jiang Zhong, Xue Li, Unified collaborative filtering model based on combination of latent features, Expert Systems with Applications, 2010;37: 5666–5672.

[7] Klasnja-Milicevic, A., Vesin, B., Ivanovic, M., Budimac, Z., e-learning Personalization Based on Hybrid Recommendation Strategy and Learning Style Identification. Computers & Education, 2011; 56(3): 885-899

[8] Khribi, M.K., Jemni, M. Nasraoui, O., Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In Advanced Learning Technologies. ICALT’08. Eighth IEEE International Conference on. 2008;241–245.

[9] Liang, G., Weining, K., Junzhou, L., Courseware recommendation in e-learning system. In Advances in web based learning, ;2006;10–24.

[10] Liu, F., Shih, B., Learning activity-based e-learning material recommendation system, In Proceedings of the ninth IEEE international symposium on multimedia workshops, 2007;343–348.

[11] Luo, J., Dong, F., Cao, J., Song, A., A context-aware personalized resource recommendation for pervasive learning, Cluster Comput, 2010;13:213–239

[12] Pazzani, M., Billsus, D., Content-based recommendation systems. The adaptive web, 2007;325–341.

[13] Robin, B. Hybrid: Recommender systems: Survey and experiments. Journal of User Modeling and User-Adapted Interaction, 2002;12,:331–370.

[14] Romero, C., Ventura, S., Zafra, A., Bra, P., Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. Computers and Education, 2009;53:828–840

[15] Salehi M. Tavakkoli-Moghaddam, R., A genetic algorithm-based grouping method for a cell formation problem with the efficacy measure I. J. of Industrial and Systems Engineering, 2010;6(3):340-359.