Journal of Discrete Mathematical Sciences and Cryptography

ISSN: 0972-0529 (Print) 2169-0065 (Online) Journal homepage: https://www.tandfonline.com/loi/tdmc20

A framework to alleviate common problems from recommender system : A case study for technical course recommendation

Poonam Singh, Sachin Ahuja, Vanita Jaitly & Shaily Jain

To cite this article: Poonam Singh, Sachin Ahuja, Vanita Jaitly & Shaily Jain (2020) A framework to alleviate common problems from recommender system : A case study for technical course recommendation, Journal of Discrete Mathematical Sciences and Cryptography, 23:2, 451-460, DOI: 10.1080/09720529.2020.1728899

To link to this article: https://doi.org/10.1080/09720529.2020.1728899

Published online: 14 May 2020.

Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 23 (2020), No. 2, pp. 451–460
DOI : 10.1080/09720529.2020.1728899

A framework to alleviate common problems from recommender system : A case study for technical course recommendation

Poonam Singh[§]
School of Computer Applications
Chitkara University
Himachal Pradesh
India

Sachin Ahuja[†]
Institute of Engineering and Technology
Chitkara University
Patiala 140401
Punjab
India

Vanita Jaitly[*]
Department of Computer Applications
Manipal University Jaipur
Jaipur 303007
Rajasthan
India

Shaily Jain[‡]
School of Engineering and Technology
Chitkara University
Himachal Pradesh
India

Abstract

Recommender systems face multiple challenges like cold-start, sparsity, first-rater and scalability. This study proposes an ontology based framework which can resolve these challenges simultaneously. The Ontology is prepared using protégé and visualized online

[§]*E-mail:* `poonam.cse@chitkarauniversity.edu.in`
[†]*E-mail:* `sachin.ahuja@chitkara.edu.in`
[*]*E-mail:* `vanita.jaitly@jaipur.manipal.edu (Corresponding Author)`
[‡]*E-mail:* `shaily.jain@chitkarauniversity.edu.in`

using Graphical Ontology Editor OWLGrEd .Ontology is trust based where degree of trust is the ratings given by users. The framework is used for technical course recommendation. To eliminate cold start, user will be provided a form with top ‘k’ factors. These top “k” factors are the significant factors obtained as a result of partial correlation analysis. The selection criterion was level of significance <0.05.The recommendation will be done on the basis of the ratings user provides to these “k” factors. Sparsity and First-rater is handled by using the Knowledge of trust from the knowledge base made from trust base ontology and provides the recommendation. The proposed framework is scalable and is implemented using offline forms. The precision rate of the proposed system is more than 95% for 3 algorithms out of total 6 implemented algorithms. The only limitation of the proposed framework is that it is domain specific because it is based on ontology.

Subject Classification: *68R01*

Keywords: *Course recommendation; Ontology based framework, Ontology based recommendation, Cold-start; First-rater, Ontology, Scalability, Sparsity*

1. Introduction

A recommendation system despite of its benefits faces various challenges. Since the recommendation is on the basis of the data available, there exist many challenges where sufficient historic data is not available for recommendation and hence leads to inaccurate recommendations. Few of shortcomings of recommender system are scalability, sparsity, first-rater and cold-start problems [1-4]. Where cold start problem is a condition when we have to give recommendation to a user which do not have any rating. In the presence of a large volume of data scalability became one of the major challenges specifically for memory based models. Sparsity is a condition when the number of ratings required to recommend a product/ item are not sufficient. First rater problem, sometimes referred as early-rater arises when we cannot recommend a new item .It is an item which is not yet checked or rated by any user.

Presence of domain knowledge in Ontology based systems removes the common problems like data sparsity and cold-start [5-7].

A lot work is done to resolve these problems individually but very few attempts are made to resolve them simultaneously [8]. In the work done we developed a framework to deal with these problems simultaneously.

2. Proposed Framework

In the proposed Ontology based framework the problem of cold-start, first-rater, scalability and sparsity will be handled. This approach is basically divided into four phases.

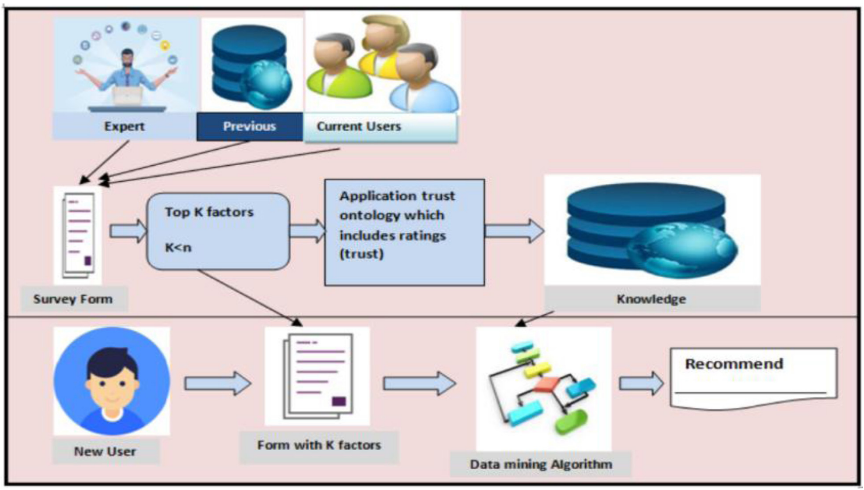

Figure 1

Ontology based framework for recommender systems

In the first phase, factors which effects choice of users from different perspective will be identified. Top ‘k’ factors will be identified in following steps:

Step 1: Design survey form to take the inputs from current users, experts and historical data let’s say the form contain ‘n’ factors.

Step 2: Apply data validation.

Step 3: Collect the data from current users (every type of user must be included). The respondents will rate the ‘n’ factors (determined in step 1)

In the second phase the factors identified in first phase will be analyzed. Find the top ‘k’ factors which affects choice of user (where $k<n$). Apply partial correlation test to find the significant factors which affect the choice of user. Select those “k” factors where level of significance is less than .05. Value of ”k” is not fixed, can vary in different context. In Third phase create the trust base application ontology where ratings of top ‘k’ factors will be the trust.

Fourth phase: - New user will be given a form (form 2) which will have the ‘k’ factors which were identified in second phase. Using the knowledge from ontology and form’s rating of user, we will recommend the user.

Figure 1 represents the proposed framework which incorporates the above steps.

Table 1

Entities and relationships for trust based ontology

Entities	Relationship between identified entities
Course	Course is offered by university
Institute	Course is offered by institute/college
Student	Course is taught by faculty
University	University employs faculty
Faculty	Institute/College employs faculty
Regulatory Committees	Institute is affiliated to University
Ranking Committees	University is accredited by regulatory committees
Family	University is ranked by ranking committees
Alumni	Student takes admission in course
Agent	Family and Friends suggest student about course
Friend	Agents advice students

3. Implementation of framework for technical course recommender system

The proposed framework is implemented on technical course recommender system. To make the trust ontology the first requirement is to identify the entities involved in course selection. Extensive literature review [9-12] and interviews of 118 students, enrolled technical courses were conducted to identify the entities. The students interviewed were enrolled in first Year of two Universities in India: Chitkara University, Punjab & Chitkara University, Himachal Pradesh. Following entities were identified to be involved in course selection along with the relationship between the entities as follows:

3.1 *Implementation of trust based Ontology*

To determine the degree of trust in trust based ontology the ratings of students were collected using Google formThe respondents were 889 students of First Year of two Universities in India: Chitkara University, Punjab & Chitkara University, Himachal Pradesh

In the Google form we included 52 factors. The students provided the ratings on likert scale 0 to 4 on the basis of increasing importance of the factor for opting the particular course.

Table 2

Average rating of each category

Category	Average Rating
Personal/social repute	1.8
University Repute	1.8
Personal/Professional Growth	1.72
Course	1.7
Institutional environment	1.6
Institutional facility	1.5
Personal needs	1.3
Financial	1.2
Recommendations	1.07

These factors were further classified into 9 categories. The categories were further ranked on the basis of the average ratings received from the students as shown in Table 2.

The ontology was prepared using Protégé 5.2.0. The Ontology is visualized using WebVOWL Available at http://vowl.visualdataweb.org/webvowl.html. The output of the protégé is Trust based ontology for course recommender system as shown in Figure 2.

Figure 2

Trust based Ontology for course recommender system

5. Results and Discussion

We applied top six prediction algorithms on the data to recommend the users. These algorithms were selected from the study [13] which found top 10 data mining algorithms on the basis of voting. All of the experiments of classification were performed on WEKA 3.8.0. For comparing the performance through experiment all of the algorithms were applied using 10 fold cross validation.

According to the comparison highest rating was of Naive-Bayes which was 7.25 followed by K nearest neighbour, AdaBoost, Classification and regression tree, Bagging and C4.5.

Naive-Bayes is an easy form of Bayesian network .Naive Bayes became popular because it have two major leads over other Bayesian networks First is simple and easy construction and second is its efficiency in the process of classification[14, 15]. Recently a new implementation of Naive Bayes classifier is proposed [16] to analyze and trim the actual size of textual log files generated by operating systems and other applications.

In K Nearest Neighbour classification k nearest neighbours are used to make a class. An analysis of K nearest Neighbour[17, 18] summarized that it is popular because its implementation is easy. Recently a new variation of K nearest neighbour is proposed [19].

Adoboost by Schapire and Freund is the boosting algorithm. Boosting algorithm is one which combines more than one weak algorithm and hence boost the performance [20, 21].

Classification and regression tree is a technique which can be executed in four steps:-tree building,stop the process of building the tree , tree pruning ,select optimum tree. The main advantage of the method is that it can deal with the missing values [22, 23].

Bagging classifier is the type of voting algorithm. Voting algorithm has two types of methods Bagging and Boosting [24, 25]. A new algorithm based on bagging is proposed [26] specifically for field of health care.

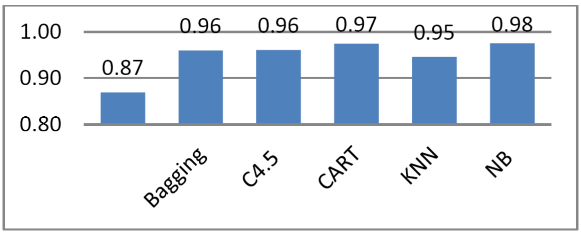

Figure 3

Precision rates of all of the algorithms

Table 3
Summary of performance of Classifiers

Algorithm	TP Rate	FP Rate	MAE
AdaBoost	0.87	0.546	0.12
Bagging	0.98	0.111	0.04
C4.5	0.98	0.139	0.03
CART	0.86	0.71	0.16
KNN	0.96	0.154	0.02
NB	0.78	0.134	0.15

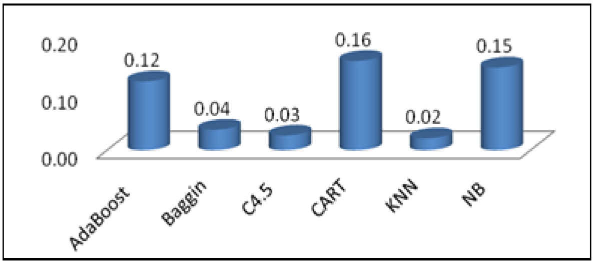

Figure 4
Mean absolute error

C4.5 was introduced by Quinlan in 1993.It is serially implemented algorithm. The algorithm can deal with both categorical as well as continuous data. In the process tree pruning leaf nodes replace the internal nodes which minimize the error rate. Splitting attribute is formulated using gain ratio impurity method [27]. The figure 3 represents the precision rate of all of the algorithms.

The bagging algorithm ,(baseline algorithm was J48) got the highest precision rate that is 98.20%.For K Nearest Neighbour where value of k was 3 Table 3 shows the summary of performance of all six classifiers. Figure 4 exhibits the Mean absolute error .This is the difference between predicted value and original value is lowest for K Nearest Neighbour.

5. Future scope

Availability of data in specific domains like online shopping, web is full of data so for online users it is easy to recommend any item or service using available data but there are many domains where users are new to that domain and there exists no historic data that can be used for recommendation e.g. domain of education students have many choices and they do not have sufficient knowledge about institute. In absence of data for such domains a well defined framework is required using which they can get recommendations. The proposed framework is able to solve cold-start, sparsity, first-rater and scalability. The precision rate of 3 algorithms are showing more than 95% and for rest of the algorithms it is more than 78%. The framework is Ontology based and hence domain specific ontology needs to be prepared. Researcher can apply trust based ontology for domain where data is insufficient to provide reliable recommendation. Researchers can make application ontology for each course which will give the better understanding of every course. This will allow reusability and can be mapped to any application.

References

[1] Moreno, N Maria Segrera, SaddysLopez, F Vivian and Munoz, Maria Dolores and Sanchez, Angel Luis (2016)," Web mining based framework for solving usual problems in recommender systems. A case study for movies' recommendation" in Neurocomputing Vol. 176, Elsevier,pp. 72—80.

[2] Zhang, Xi and Cheng, JianQiu, Shuang Zhu, GuiboLu, Hanqing (2015) , "DualDS: A dual discriminative rating elicitation framework for cold start recommendation" in Knowledge-Based Systems,Vol. 73, Elsevier,pp. 161-172

[3] Pirasteh, Parivash Hwang, DosamJung, Jason J (2015),"Exploiting matrix factorization to asymmetric user similarities in recommendation systems "in Knowledge-Based Systems, Vol. 83, Elsevier,pp. 51-57

[4] GediminasAdomavicius and JingjingZhang(2015),"Improving stability of recommender systems: A meta-algorithmic approach", in IEEE Transactions on Knowledge and Data Engineering, Vol. 27, IEEE, pp. 1573-1587

[5] Zhao, X., Niu, Z., Wang, K., Niu, K., & Liu, Z. (2015). Improving top-N recommendation performance using missing data. *Mathematical Problems in Engineering*, 2015.

[6] Morente-Molinera, J. A., Kou, G., González-Crespo, R., Corchado, J. M., & Herrera-Viedma, E. (2017). Solving multi-criteria group decision making problems under environments with a high number of

alternatives using fuzzy ontologies and multi-granular linguistic modelling methods. Knowledge-Based Systems, 137, 54-64.

[7] Makwana, K., Patel, J., & Shah, P. (2017, March). An Ontology Based Recommender System to Mitigate the Cold Start Problem in Personalized Web Search. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 120-127). Springer, Cham.

[8] Singh, P., Ahuja, S. and Jain, S., 2019. Latest Trends in Recommender Systems 2017. In Advances in Data and Information Sciences (pp. 197-210). Springer, Singapore.

[9] Kallio, R. E. (1995). Factors influencing the college choice decisions of graduate students. Research in Higher Education, 36(1), 109-124.

[10] DesJardins, S. L., Dundar, H., & Hendel, D. D. (1999). Modeling the college application decision process in a land-grant university. Economics of Education Review, 18(1), 117-132.

[11] Ahmad, S. Z., & Hussain, M. (2017). An investigation of the factors determining student destination choice for higher education in the United Arab Emirates. Studies in Higher Education, 42(7), 1324-1343.

[12] Pugh, D. L. (2017). Factors Affecting African-American Enrollment and Intent to Enroll in an Advanced Placement Program in a Suburban High School.

[13] Settouti, N., Bechar, M. E. A., & Chikh, M. A. (2016). Statistical comparisons of the top 10 algorithms in data mining for classification task. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), 46-51

[14] Cheng, J., & Greiner, R. (1999, July). Comparing Bayesian network classifiers. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (pp. 101-108). Morgan Kaufmann Publishers Inc.

[15] Albayati, M., & Issac, B. (2015). Analysis of intelligent classifiers and enhancing the detection accuracy for intrusion detection system. *International Journal of Computational Intelligence Systems*, *8*(5), 841-853.

[16] Ritchey, R. P., Shearer, G. G., & Renard, K. D. (2019). Nave Bayes Log File Reduction and Analysis (No. ARL-TR-8624). US Army Research Laboratory Aberdeen Proving Ground United States.

[17] Cunningham, P., & Delany, S. J. (2007). *k*-Nearest neighbour classifiers. Multiple Classifier Systems, 34(8), 1-17.

[18] Lee, I., Kwak, M., & Han, D. (2016). A dynamic *k*-nearest neighbor method for WLAN-based positioning systems. *Journal of Computer Information Systems*, 56(4), 295-300.

[19] Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient knn classification with different numbers of nearest neighbors. *IEEE transactions on neural networks and learning systems*, 29(5), 1774-1785.

[20] Schapire, R. E. (2013). Explaining adaboost. In Empirical inference (pp. 37-52). Springer, Berlin, Heidelberg.

[21] Zhang, P. B., & Yang, Z. X. (2016). A novel adaboost framework with robust threshold and structural optimization. *IEEE transactions on cybernetics*, 48(1), 64-76.

[22] Lewis, R. J. (2000, May). An introduction to classification and regression tree (CART) analysis. In Annual meeting of the society for academic emergency medicine in San Francisco, California (Vol. 14).

[23] Van Aardt, J. A. N., & Norris-Rogers, M. (2008). Spectral–age interactions in managed, even-aged Eucalyptus plantations: application of discriminant analysis and classification and regression trees approaches to hyperspectral data. *International Journal of Remote Sensing*, 29(6), 1841-1845.

[24] Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine learning, 36(1-2), 105-139.

[25] Puri, Vikram, Anand Nayyar, and Linesh Raja. "Agriculture drones: A modern breakthrough in precision agriculture." *Journal of Statistics and Management Systems* 20.4 (2017): 507-518.

[26] Lee, S. J., Xu, Z., Li, T., & Yang, Y. (2018). A novel bagging C4. 5 algorithm based on wrapper feature selection for supporting wise clinical decision making. *Journal of biomedical informatics*, 78, 144-155.

[27] Ma, B. L. W. H. Y., Liu, B., & Hsu, Y. (1998, August). Integrating classification and association rule mining. In Proceedings of the fourth international conference on knowledge discovery and data mining (pp. 24-25).