DISADVANTAGES and RESOLUTION of applying association rule mining in learning management systems

Meenakshi

Asst. Prof., Dept. of CSE, GGIAET, Gurgaon, Haryana

ABSTRACT

In this paper, investigation will be carried out on the application of association rule mining in learning management systems and e-learning systems. We shall explain the precise information discovery process, its main disadvantages and few possible resolutions for the same.

Keywords: association rule mining, e-learning, information discovery

I. INTRODUCTION

These days, Colleges, schools, universities and also individual persons are installing Learning Management Systems (LMS) rapidly as they want to add web technology to their courses and also to complement customary face-to-face courses [1]. LMS systems collect a vast amount of information and that is very precious for analyzing the students' performance and could create an excellent compilation of educational data [2]. LMS systems can record all the student activities involved. For instance: Writing, Reading, various task performance and internal communication.

Data generated in this manner is huge and it becomes very tedious to analyze this data manually. Special capable approach towards this analysis objective is the usage of DMT; data mining techniques.

KDD i.e. discovery in databases is nothing but the automatic extraction of inherent and remarkable patterns from massive data collection [3]. Association rules mining is one of the mainly well considered data mining tasks. It produces if-then statements

Concerning attribute-values and discovers interaction among attributes in databases [4]. An association rule X ⇒ Y expresses that if X occurs then probability of occurring Y is high as well. X is called antecedent and Y is called consequent of the rule. A consequent is an item that is found in combination with the antecedent. It is measured by its support and confidence. **Support** is a sign of how frequently the items appear in the database. *Confidence* indicates how many times the if/then statements found to be true. Association rule mining can be applied to e-learning.

It discovers interesting relationships from student's usage information for providing feedback to course author [6]. To find students' mistakes which occurs together usually [8], finding out the relationships between each pattern of learner's behavior [7]. For best fitting transfer model guiding the search of student learning [9], by shaping the content of most concern to the user optimize the content of an e-learning portal [10], interpreting on-line course behavior and extracting valuable patterns for helping educators web masters and [5], and personalizing e-learning based on aggregate usage profiles and a domain ontology [11]. Finally, for e-learning classification association rule mining has been used [14].

Association rule mining also used for learning of sequential patterns mining, which is a restrictive form of association rule. It concerned for finding relevant patterns between data examples where the values are delivered in a sequence.[1] It is generally supposed that the values are distinct, and thus time series mining is closely related, but frequently considered a different action. SPM is a case of structured data mining. For determining the learners' activities extraction of sequential patterns are used in e-learning and can also be used in customizing and adapting resource delivery [12]; To discover and compare with anticipated behavioral patterns particular by the mentor that explain an ideal learning path [13].

II.THE ASSOCIATION RULE MINING PROCESS IN LMS

Collecting data, preprocessing, applying the actual data mining tasks and post-processing are the general KDD process [15] next steps. These steps are particularized for association rule mining in the LMS domain.

• **Data Collection.** Current LMSs do not store logs as text files. Usually a relational database is used that stores all the systems information. Academic results, Data of user interaction, personal profile etc Databases are more powerful, flexible and bug-prone than the normally textual log files for assembly detailed access and high level practice. The LMSs keep detailed logs of all activities that students perform. Not only every click that students make for navigational purposes (low level information) is stored, but it also stores elapsed time, test scores etc. (high level information).

• **Data pre-processing.** Almost all of the conventional data pre-processing tasks, such as user identification, cleaning of data, identification of session, transaction identification, enrichment and data transformation and data reduction are not necessary in LMS. LMS Data's pre-processing is very simple because LMS store the data for analysis purposes, in comparison to the usually observational datasets in data mining, that were generated to maintain the operational setting and not for analysis in the first place. Database and user authentication (password protection) are also used in LMS which permits identifying the users in the logs.

• **Applying the mining algorithms.** In this step it is compulsory: **A**) to choose he exact association rule mining algorithm and execution; B) to constitute the parameters of the algorithm, like support and confidence threshold and others; **C**) to recognize data file or table will be used for the mining; **D**) and to identify a few other limitations, such as the maximum number of items and what specific attributes can be present in the predecessor or resultant of the discovered rules.

• **Data post-processing.** The received or achieved results or rules are interpreted, evaluated and used by the tutors and guides for advance actions. The final objective is to position the results into usage. The discovered information is used (in form of if-then rules) for making decisions by teachers about the students and the LMS activities of the classes and lessons in order

to improve the students' learning. So, data mining algorithms have to articulate the output in a intelligible format.

III. DISADVANTAGES AND RESOLUTIONS

Most of the research efforts went in the first place to improving the algorithmic performance [16] in the association rule mining area, and in the second place into dropping the output set by permitting the likelihood to express constraints on the preferred results.

From last few years, a various algorithms have been developed that address these issues through the fine-tuning of search strategies, pruning techniques and data structures. Most algorithms center on the unequivocal discovery of all rules that suit minimal support and confidence limitations for a given dataset, increasing consideration is being given to specialized algorithms that attempt to improve processing time or facilitate user interpretation by reducing the result set size and by incorporating domain knowledge [17].

There are ample of other precise problems linked to the application of association rule mining from e-learning data. While trying to solve these problems, we must consider the purpose of the association models and the data they come from.

Presently, data mining tools are designed more for power and flexibility rather than for simplicity. Most of the current data mining tools are too complex for an educator and teacher to use and their features go well beyond the scope of what an educator might require.

It is most enviable that teachers participate directly in the iterative mining process so that more valuable rules can be obtained. Some of the main drawbacks of association rule algorithms in e-learning are: the algorithms have too many parameters for some in experienced person in data mining and the obtained rules are in abundance, most of them non-interesting and with low unambiguousness. In the following subsections, we shall face off these issues.

3.1 Finding the appropriate parameter settings of the mining algorithm

Before Execution, association rule mining algorithms need to be configured. So, the user has to give suitable values for the parameters well in advance (often leading to plenty or very few rules) in order to achieve a respective number of rules. A relative study between the main algorithms that is at present used to discover association rules can be found in [18]: Apriori [19], FP-Growth [20], Magnum Opus [21], and Closet [22]. Most of these algorithms require the user to set two thresholds, the minimal confidence and minimal Support and find all the rules that exceed the thresholds specified by the user. Therefore, the user must possess a definite amount of skill in order to find the correct settings for support and confidence to obtain the most excellent rules. One possible solution to this problem can be the use of a parameter-free algorithm or with less parameter. For instance, the Weka [23] package implements an Apriori type algorithm that resolves this problem on practical basis.

One more enhanced version of the Apriori algorithm is the Predictive Apriori algorithm [24], which by design resolves the problem of balance between these two parameters, maximizing the probability of making an accurate prediction for the data set. In order to attain this, a factor called the exact expected predictive accuracy is defined and calculated using the Bayesian method, that provides information about the accuracy of the rule found. In this way the client only has to identify the maximal number or rules to discover.

IV.CONCLUSIONS

Total integration of association rule mining in e-learning systems is still in premature stage and not many real and fully operative implementations are available. In this paper, outlining of few of the main drawbacks for the application of association rule mining in learning management systems has been attempted and few possible solutions for problems have been detailed.

REFERENCES

1. Rice, W.H.: Moodle E-learning Course Development. A complete guide to successful learning using Moodle. Packt publishing (2006).

2. Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C.: An educational data mining tool to browse tutor-student interactions: Time will tell! In: Proc. of the Workshop on Educational Data Mining (2005) 15–22.

3. Klosgen, W., & Zytkow, J.: Handbook of data mining and knowledge discovery. Oxford University Press, New York (2002).

Drawbacks and solutions of applying association rule mining 23

4. Agrawal R., Imielinski, T., Swami, A.N.: Mining Association Rules between Sets of Items in Large Databases. In: Proc. of SIGMOD (1993) 207-216.

5. Zaïane, O.: Building a Recommender Agent for e-Learning Systems. In: Proc. of the Int. Conf. in Education (2002) 55-59.

1-8.

6. Romero, C., Ventura, S., Bra, P. D.: Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 14:5 (2004) 425–464.

7. Yu, P., Own, C., Lin, L.: On learning behavior analysis of web based interactive environment. In: Proc. of the Int. Conf. on Implementing Curricular Change in Engineering Education (2001) 1-10.

8. Merceron, A., & Yacef, K.: Mining student data captured from a web-based tutoring tool. Journal of Interactive Learning Research, 15:4 (2004) 319–346.

9. Freyberger, J., Heffernan, N., Ruiz, C.: Using association rules to guide a search for best fitting transfer models of student learning. In: Workshop on Analyzing Student-Tutor Interactions Logs to Improve Educational Outcomes at ITS Conference (2004) 1-10.

10. Ramli, A.A.: Web usage mining using apriori algorithm: UUM learning care portal case. In: Proc. of the Int. Conf. on Knowledge Management (2005) 1-19.

11. Markellou, P., Mousourouli, I., Spiros, S., Tsakalidis, A.: Using Semantic Web Mining Technologies for Personalized e-Learning Experiences. In: Proc. of the Int. Conference on Web-based Education (2005) 1-10.

12. Zaïane, O., Luo, J.: Web usage mining for a better web-based learning environment. In: Proc. of Int. Conf. on advanced technology for education (2001) 60–64.

13. Pahl, C., Donnellan, C.: Data mining technology for the evaluation of web-based teaching and learning systems. In: Proc.of Int. Conf. E-learning (2003) 1-7.

14. Castro, F., Velllido, A. , Nebot, A. and Mugica, F.: Applying Data Mining Techniques to e-Learning Problems: a Survey and State of the Art. Evolution of Teaching and Learning Paradigms in Intelligent Environment. Springer (2007) 183-221.

15. Hamalainen, W., Vinni, M.: Comparison of machine learning methods for intelligent tutoring systems. In: Proc. of Int. Conf. in Intelligent Tutoring Systems (2006) 525–534.

16. Ceglar, A., Roddick, J.F.: Association mining. ACM Computing Surveys, 38:2 2006 1-42. 30. Goethals B., Nijssen S., Zaki, M.: Open source data mining: workshop report. SIGKDD Explorations, 7:2 (2005) 143-144.

17. Zheng Z., Kohavi R., Mason L. Real world performance of association rules. In: Proc. Of the Sixth ACM-SIGKDD (2001) 86-98.

18. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc. of Int. Conf. on Very Large Data Bases (1996) 487-499.

19. Han, J., Pei, J., and Yin, Y.: Mining frequent patterns without candidate generation. In: Proc. of ACM-SIGMOD Int. Conf. on Management of Data (1999) 1-12.

20. Webb, G.I.: OPUS: An efficient admissible algorithm for unordered search. Journal of Artificial Intelligence Research 3 (1995) 431-465.

21. Pei, J., Han, J., Mao, R.: CLOSET: An efficient algorithm for mining frequent closed itemsets. In Proc. of ACM_SIGMOD Int. DMKD (2000) 21-30.

22. Weka. (2007) Available at http://www.cs.waikato.ac.nz/ml/weka/.

23. Tobias S.: Finding Association Rules That Trade Support Optimally against Confidence.In: Proc. of the European Conf. of PKDD (2001) 424-437.