Pros and Cons of Applying Association Rule Mining in LMS

Preeti Gupta, Ritesh Rastogi

Department of MCA, Noida Institute of engineering and technology India

Abstract: ***We present the unique expertise discovery technique, its primary shortcomings, and some potential solutions to clear them up in this study paper, which surveys the applications of association rule mining in e-learning knowledge of structures, notably getting to know control systems*.**

Keyword: LMS, Association Rule Mining, Data mining

Introduction

Learning management systems (LMS) have become increasingly popular in recent years, with universities, community institutions, colleges, businesses, and even character teachers using them to add internet technology to their guides and augment traditional face-to-face courses [1]. LMS structures collect a large amount of data that is quite useful for analyzing students' behavior and will result in a gold mine of instructional statistics [2]. They can chronicle everything related to scholar reports, such as studying, writing, taking tests, doing other activities, or even chatting with pals. However, manually researching this data is quite difficult due to the massive volumes of records that those systems can generate on a regular basis. Association rule mining is a very promising method for achieving this analysis goal which uses the statistics mining strategies.

The automatic extraction of implicit and fascinating styles from large fact sets is known as data mining or knowledge discovery in databases (KDD) [3]. One of the most thoroughly researched records mining tasks is association guidelines mining. It finds associations between properties in databases and generates if-then sentences about attribute values [4]. An association rule X Y states that if X occurs in a database transaction, there is a good likelihood that Y will occur as well. The antecedent and consequent of the rule of thumb are called as X and Y, respectively. The strength of a law like this is determined by how much support and confidence it has. The fraction of interactions with X in the database that have the guideline's self-assurance the resultant Y additionally. The assist of the rule of thumb is the proportion of transactions inside the database that include each the antecedent and the ensuing.

Association rule mining has been implemented to e-learning structures for traditionally association analysis (finding correlations among items in a dataset), conjunction with, for example, the following tasks: constructing recommender agents for learning sports or shortcuts [5], automatically guiding the learner's sports and intelligently generate and endorse mastering substances [6], identifying attributes characterizing patterns of performance disparity among various groups of college students [7], uncovering interesting relationships from student usage records in order to provide feedback to the course creator [8,] determining the relationships among each sample [12], extracting useful styles to help educators and internet masters evaluating and deciphering route activities [5], and personalizing e-learning primarily based on aggregate utilization profiles and a domain ontology [13].

Association rule mining is a restricted form of association rule mining in that it considers not only the occurrences themselves, but also the order in which the occurrences of the items occur. The extraction of sequential styles has been used in e-learning for evaluating beginners' sports and may be used in adapting and customizing help delivery [14]; discovering and comparing with expected behavioral patterns specified by way of the instructor that describes

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

an excellent studying path [15]; giving an indication of how to nicely arrange the academic net space and be able to make suggestions to rookies who percentage comparable characteristic to extraordinary agencies of learners [17]; helping the assessment and validation of mastering web page designs [18]; figuring out interplay sequences indicative of troubles and styles which are markers of fulfillment [19].

Finally, affiliation rule mining was applied to e-learning for class [20]. The primary difference between classification rules and regular affiliation rules is that classification rules have an inescapable condition in the result, which is the elegance identification call. They were used in learning about fabric organization [21], student learning tests [22, 23, 24], leadership model for students' conduct [25, 26], and evaluation of educational web sites [27].

This paper is written in the following format: The KDD method for affiliation rule mining in e-studying is described in section 2. The main disadvantages and solutions to using association rule algorithms in LMS are discussed in Section 3. Finally, in part four, the conclusions and research findings are discussed.

The association rule mining process in LMS

The overall KDD [28] has the following steps: gathering statistics, preprocessing, making use of the actual facts mining duties and publish-processing. We particularize these steps for association rule mining in the LMS area.

• Collecting data. The majority of today's LMSs do not keep logs as textual content documents. Alternatively, they frequently employ a relational database to store all of the structure records, such as the users' non-public data (profile), instructional outcomes, and user interaction data. For obtaining unique access and large levels of utilization, databases are more effective, flexible, and virus-resistant than normally textual log entries. The disadvantages and solutions of using the affiliation rule to mine 17 information from all of the services available in the LMS. The LMSs keep detailed records of all sports that kids participate in. Each click on that pupils make for navigational functions is ineffective. (Low degree data) is saved, but also checks ratings, elapsed time, and so forth(high stage information).

• Data pre-processing. The majority of today's LMSs do not keep logs as textual content documents. Alternatively, they frequently employ a relational database to hold all of the structure records, such as user public data (profile), instructional outcomes, user interaction data, and so on. For obtaining unique access and large levels of utilization, databases are more effective, flexible, and virus-resistant than normally textual log entries. The disadvantages and solutions of using the affiliation rule to mine 17 information from all of the services available in the LMS. The LMSs keep detailed records of all sports that kids participate in. Each click that students make for navigational functions (low degree data) is kept, but it also checks ratings, elapsed time, and other factors. (High level information).

• Applying the mining algorithms. . This step is critical: 1) to choose the precise association rule mining set of rules and implementation; 2) to configure the parameters of the algorithm, such as assist and confidence thresholds and others; three) to determine which desk or information file will be used for the mining; and four) to specify some other restrictions, such as the maximum number of items and what specific attributes may be given inside the antecedent or consequence of the discovered.

• Data post-processing. The trainer interprets, evaluates, and applies the obtained outcomes or rules to future moves. The very final goal is to put the effects into action. Teachers employ the found data (in the form of if-then rules) to make decisions about the scholars and the direction's LMS activities in order to improve the scholars' learning. As a result, information mining algorithms must provide the output in an intelligible way by employing standardized e-getting to know metadata, for example. It's critical to note that standard academic information units are frequently small [28]. When compared to databases used in various information mining industries, such as ecommerce program with thousands of clients.

Furthermore, depending on how much data the LMS stores about each student's interaction with the system, the aggregate range of times or transactions can be extremely large (and at what ranges of granularity). As a result, the range of available times is far greater than the number of pupils. And, as previously stated, educational records have one benefit over many other domain names [28]: the facts units are often extremely simple, i.e., the values are correct and do not include any noise from measurement devices.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Drawbacks and solutions

The majority of research efforts in the affiliation rule mining area went towards two areas: enhancing algorithmic overall performance [29] and decreasing the output set by enabling the ability to specify limits at favored outcomes.Over the last decade, a number of algorithms have been developed to solve these issues by improving search tactics, pruning techniques, and record structures. specialized algorithms that try to improve processing time or facilitate consumer interpretation by lowering the result set length and incorporating area information are getting more attention at the same time as maximum algorithms focus on the explicit discovery of all rules that fulfill minimal support and self assurance constraints for a given dataset [30].There are also some particular issues with the utility of mining affiliation rules from e-studying data. When trying to solve these issues, it's important to remember the objective of the association models and the data they use.

These days, most data mining equipment is built for strength and versatility rather than simplicity. Most existing data mining tools are too complicated for educators to use, and their functions go much beyond what an educator could need.

As a result, the courses administrator is more likely to use data mining strategies to generate reports for teachers, who then use these findings to make judgments about how to improve student learning and web courses. However, instructors who participate directly in the iterative mining process are more likely to achieve more valuable policies. However, most instructors just use the feedback provided by the collected regulations to make decisions about course modifications, such as finding activities or students with problems, and so on.

Some of the major disadvantages of association rule algorithms in e-mastering include: the utilized algorithms contain much too many parameters for any non-expert in information mining, and the generated policies are considerably too numerous, with the majority of them being uninteresting and difficult to comprehend. We can address such issues in the subsections that follow.

Finding the appropriate parameter settings of the mining algorithm

Before they can be used, association rule mining algorithms must be configured. If you want a good range of rules, the consumer must submit adequate values for the parameters in advance (which frequently leads to too many or too few recommendations). A comparison of the main techniques that are currently utilised to discover association policies may be seen in [31]: Apriori [32], FP-growth [33], MagnumOpus [34], Closet[35]. Most of these algorithms ask the user to choose two thresholds, a minimum guide and a minimum self-confidence, and then search for all regulations that surpass the consumer's specified thresholds. As a result, the customer must have a certain level of knowledge in order to discover the proper settings for guidance and the confidence to obtain the best instructions.

One viable solution to this problem is to use a parameter-light or parameter-less method. The Weka [36] bundle, for example, implements an Apriori-style set of rules that partially overcomes this problem. This set of rules reduces the minimal aid iteratively, using a component delta help (s) provided by the person, until the minimum guide is reached or a required variety of policies (NR) is generated.

The Predictive Apriori algorithm [37] is another modified model of the Apriori algorithm that robotically addresses the problem of stability between those two parameters, maximising the chance of making an accurate forecast for the records set. In order to reap this benefit, a parameter known as the precise projected predictive accuracy is created and calculated using the Bayesian technique, which provides data on the correctness of the observed guideline. In this case, the person merely needs to provide the maximum number of options or guidelines to investigate[38].With the help of comparing the two prior algorithms, experimental assessments were carried out on a Moodle path. The final results confirmed that Predictive Apriori outperformed Apriori-type set of rules when the s component was used.

Discovering too many rules

Traditional association algorithms may have simple and efficient applications. However, association rule mining algorithms frequently uncover a large number of policies and cannot guarantee that all of them are meaningful. The guidance and confidence elements can be utilised to find intriguing rules with values higher than a threshold price

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

for those characteristics. Although those options allow for the trimming of a large number of institutions, another common constraint is to specify the attributes that should or cannot be given as a result of the located criteria.

Another option is to evaluate and submit-prune the existing regulations, allowing you to choose the most fascinating policies for a certain issue. The use of objective interestingness metrics has generally been discouraged [39], including the previously mentioned support and confidence, as well as additional measures such as Laplace, chi-rectangular statistic, correlation coefficient, entropy benefit, gain, hobby, conviction, and so on. These measures can be used to rank the received regulations so that the individual can choose the ones with the highest values in the measures in which he or she is most interested.

Subjective measures, or measures that are dependent on subjective elements controlled by the user, have grown increasingly important [40]. Most subjective tactics include consumer participation, which will specify which policies are of interest based on his or her prior knowledge. The following are some recommended subjective measures [41]:

• Unexpectedness: Rules are interesting if they are unknown to the user or contradict the user's knowledge.

• Actionability: Policies are thrilling if customers can do some thing with them to their gain.

The multiplicity of rules can be reduced by displaying just unexpected and actionable rules to the teacher rather than all of the discovered rules [38]. [41] proposes an Interestingness Analysis Device (IAS). It contrasts restrictions with the person's knowledge in the hobby's field. Let U be the set of user specifications that represents his or her knowledge domain. This method, given a collection of determined association policies, uses a pruning approach to eliminate redundant or irrelevant policies by rating and sorting them into four categories:

• Conforming rules: a observed rule $Ai \in A$ conforms to a bit of consumer's information Uj if each the antecedent and the ensuing parts of Ai fit the ones of $Uj \in U$ well.

• Unexpected consequent rules: a determined rule $Ai \in A$ has surprising consequents with respect to $Uj \in U$ if the antecedent part of Ai suits that of Uj nicely.

• Unexpected condition rules: a determined rule $Ai \in A$ has sudden conditions with appreciate to $Uj \in U$ if the ensuing a part of Ai fits that of Uj well, but no longer the antecedent component.

• Both-side unexpected rules: a located rule $Ai \in A$ is both-facet unexpected with recognize to $Uj \in U$ if the antecedent and consequent parts of Ai don't fit those of Uj nicely.

The guidelines are rated based on the club levels in each of these four classifications. They provide their information about the problem in query, using their own specification language, via relationships between many of the database's fields or objects.

Finally, we can use the information database, which is the most commonly used repository, to conduct subjective analysis of the policies observed [38]. The trainer could download the relevant understanding database, according to his or her profile, before running the affiliation rule mining set of rules. The filtering characteristics linked with the type of route to be studied, such as the proximity of information, the level of training, the complexity of the course, and many others, are used to personalise the results. The rules repository is generated collaboratively on the server, with professionals voting on each rule inside the repository based on instructional concerns and their experience with other similar e-learning publications.

Discovery of poorly understandable rules

The comprehensibility of the extracted recommendations is a crucial factor to consider when determining their quality. Despite the fact that the primary motivation for rule extraction is to obtain an understandable description of

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

the underlying version's speculation, this aspect of rule nice goes unnoticed due to the subjective nature of comprehensibility, which cannot be measured independently of the person using the device [42]. The character's prior experience and knowledge of the place play a vital influence in determining the comprehensibility. This contrasts with accuracy, which can be considered a part of the regulations and assessed independently of the customers.

There are a few traditional ways that have been employed to improve the understandability of found guidelines. For example, we can reduce the size of policies by limiting the number of objects allowed in the antecedent or consequence of the policy. The size of the rule has an effect on its simplicity, so the shorter the rule, the more understandable it will be. A discretization of numerical quantities is used in another way. Discretization [43] splits numerical data into categories training that is easier for the teacher to understand (specific values are greater person-pleasant for the teacher than unique magnitudes and stages).

Another way to improve the regulations' comprehension is to include area understanding and semantics, as well as to use a common and well-known terminology for the instructor. We will discover a few similar qualities to a ramification of e-learning structures such as LMS and adaptive hypermedia publishes in the context of net-based fully instructional systems. As shown in table 1, certain characteristics can be found in numerous sections or stages of the course. A unit could be a game, a lecture, a workout, or a collaborative resource, for example.

Using preferred metadata for e-learning [44] in this context facilitates the creation and maintenance of a shared knowledge base with a common vocabulary that is conducive to sharing among diverse communities of instructors. For example, the SCORM [45] standard specifies a content aggregation and monitoring version for reusable learning objects. SCORM provides a foundation for the representation and processing of metadata, but it lacks the support required for other, more particular educational monitoring, such as the use of collaborative resources. We show a proposed SCORM-based Ontology for association rule mining in LMS using current e-studying attributes in parent Other features relating to collaborative studying, in addition to the conventional SCORM attributes. This will be an excellent spot to start when it comes to repurposing information and replacing outcomes in unique mining frameworks. It will be very useful to give an ontology that explains the particular area in order to improve the comprehensibility and appropriateness of the rules [44].

As a result, the trainer will have a better understanding of the policies that incorporate notions related to the domain under investigation, such as "if fulfilment in subject matter A, then fulfilment in subject B." Finally, another idea is to use domain-specific interactive statistics mining [46], in which the individual is involved in the discovery system to find the most exciting findings repeatedly.

The mining system additionally includes domain and problem-specific representations. The person isn't only comparing the output of an automated data mining process; she's also actively involved in the creation of a new representation and the search for samples.

Conclusions and future trends

Its miles nevertheless the early days for the entire integration of association rule mining in e-mastering structures and no longer many actual and fully operative implementations are available. On this paper, we have mentioned some of the primary drawbacks for the software of affiliation rule mining in mastering management structures and we've got described a few possible answers for each hassle.

We believe that a few future research lines will focus on: developing association rule mining equipment that can be used more easily by educators; including new specific measures of interest with the inclusion of area know-how and semantic; embedding and integrating mining gear into LMSs so that teachers can use the same interface to create/preserve guides and to perform the mining procedure/obtain direct feedback/make changes in the co designing iterative, interactive, or guided mining to assist educators in applying KDD procedures, or even building an automated mining device that could execute mine robotically and unattended, so that the instructor only has to apply the offered pointers on the way to increase students' learning.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

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.

[6] Lu, J.: Personalized e-learning material recommender system. In: Proc. of the Int. Conf. on Information Technology for Application (2004) 374–379.

[7] Minaei-Bidgoli, B., Tan, P., Punch, W.: Mining interesting contrast rules for a web-based educational system. In: Proc. of the Int. Conf. on Machine Learning Applications (2004) 1-8.

[8] 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. Samidha Garg, S.N.Singh; 'Performance Comparison Of High Frequency Single Phase PWM Inverters Using PSpice',Volume No.1,Issue No.1,2013,PP.001-008,ISSN :2229-5828

[9] Satyendra Sharma, Dr.Brahmjit Singh; 'Planning and Optimization of UMTS Network Radio Link with Limited Interference',Volume No.2,Issue No.1,2013,PP.009-012,ISSN :2229-5828

[10] L.NavinkumarRao, S Gairola; 'Steady -State and Dynamic Performance of the Static Var Compensator (SVC) Phasor Model',Volume No.2,Issue No.1,2013,PP.013-016,ISSN :2229-5828

[11] Rishi Singhal, S.P. Singh and Dr V. K. Pandey; 'Design of Phase Compen-sation Circuits For Multistage OP-AMP',Volume No.2,Issue No.1,2013,PP.017-019,ISSN :2229-5828

[12] Mohd Yunus Khan, Saroj Kumar Padhi and R.K. Singh; 'Extraction ,Characterization And Engine Performance of Jatropha Curcas Oil',Volume No.2,Issue No.1,2013,PP.021-028,ISSN :2229-5828

[13] Srivastav, Vivek; 'Analytical Study of Traffic Behaviour in a Network on Chip Architecture',Volume No.2,Issue No.1,2013,PP.029-032,ISSN :2229-5828

[14] Saumya Priya Basu , Saumya Das , Sanjita Das, Manas K Das; 'Effect of Methanoic Extract of the Leaves of Calotropis Gigntea R.BR. On Leukocyte and Neutrophil Migration',Volume No.2,Issue No.1,2013,PP.033-036,ISSN :2229-5828

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.