Drawbacks and solutions of applying association rule mining in learning management systems

Enrique García[1], Cristóbal Romero[1], Sebastián Ventura[1], Toon Calders[2]

[1]Córdoba University, Campus Universitario de Rabanales, 14071, Córdoba, Spain
{egsalcines,cromero,sventura}@uco.es
[2]Eindhoven University of Technology (TU/e), PO Box 513, Eindhoven, The Netherlands
toon.calders@ua.ac.be

Abstract. In this paper, we survey the application of association rule mining in e-learning systems, and especially, learning management systems. We describe the specific knowledge discovery process, its mains drawbacks and some possible solutions to resolve them.

1 Introduction

Nowadays, Learning Management Systems (LMS) are being installed more and more by universities, community colleges, schools, businesses, and even individual instructors in order to add web technology to their courses and to supplement traditional face-to-face courses [1]. LMS systems accumulate a vast amount of information which is very valuable for analyzing the students' behavior and could create a gold mine of educational data [2]. They can record whatever student activities it involves, such as reading, writing, taking tests, performing various tasks, and even communicating with peers. However, due to the vast quantities of data these systems can generate daily, it is very difficult to analyze this data manually. A very promising approach towards this analysis objective is the use of data mining techniques.

Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections [3]. Association rules mining is one of the most well studied data mining tasks. It discovers relationships among attributes in databases, producing if-then statements concerning attribute-values [4]. An association rule $X \Rightarrow Y$ expresses that in those transactions in the database where X occurs; there is a high probability of having Y as well. X and Y are called respectively the antecedent and consequent of the rule. The strength of such a rule is measured by its support and confidence. The confidence of the rule is the percentage of transactions with X in the database that contain the consequent Y also. The support of the rule is the percentage of transactions in the database that contain both the antecedent and the consequent.

Association rule mining has been applied to e-learning systems for traditionally association analysis (finding correlations between items in a dataset), including, e.g., the following tasks: building recommender agents for on-line learning activities or

shortcuts [5], automatically guiding the learner's activities and intelligently generate and recommend learning materials [6], identifying attributes characterizing patterns of performance disparity between various groups of students [7], discovering interesting relationships from student's usage information in order to provide feedback to course author [8], finding out the relationships between each pattern of learner's behavior [9], finding students' mistakes that are often occurring together [10], guiding the search for best fitting transfer model of student learning [11], optimizing the content of an e-learning portal by determining the content of most interest to the user [12], extracting useful patterns to help educators and web masters evaluating and interpreting on-line course activities [5], and personalizing e-learning based on aggregate usage profiles and a domain ontology [13].

Association rule mining also has been applied to the learning of sequential patterns mining, which is a restrictive form of association rule mining in the sense that not only the occurrences themselves, but also the order between the occurrences of the items is taken into account. The extraction of sequential patterns has been used in e-learning for evaluating the learners' activities and can be used in adapting and customizing resource delivery [14]; discovering and comparison with expected behavioral patterns specified by the instructor that describes an ideal learning path [15]; giving an indication of how to best organize the educational web space and be able to make suggestions to learners who share similar characteristics [16]; generating personalized activities to different groups of learners [17]; supporting the evaluation and validation of learning site designs [18]; identifying interaction sequences indicative of problems and patterns that are markers of success [19].

Finally, association rule mining has been used in the e-learning for classification [20]. From a syntactic point of view, the main difference to general association rules is that classification rules have a single condition in the consequent which is the class identifier name. They have been applied in learning material organization [21], student learning assessments [22, 23, 24], course adaptation to the students' behavior [25, 26] and evaluation of educational web sites [27].

This paper is organized in the following way: Section 2 describes the KDD process for association rule mining in e-learning. Section 3 describes the main drawbacks and solutions of applying association rule algorithms in LMS. Finally, in section 4, the conclusions and further research are outlined.

2   The association rule mining process in LMS

The general KDD process [28] has the next steps: collecting data, preprocessing, applying the actual data mining tasks and post-processing. We particularize these steps for association rule mining in the LMS domain.

- **Collecting data.** Most of the current LMSs do not store logs as text files. Instead, they normally use a relational database that stores all the systems information: personal information of the users (profile), academic results, the user's interaction data, etc.. Databases are more powerful, flexible and bug-prone than the typically textual log files for gathering detailed access and high level usage

information from all the services available in the LMS. 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 also test scores, elapsed time, etc. (high level information).

- **Data pre-processing.** Most of the traditional data pre-processing tasks, such as data cleaning, user identification, session identification, transaction identification, data transformation and enrichment, data integration and data reduction are not necessary in LMS. Data pre-processing of LMS data is simpler due to the fact that most LMS store the data for analysis purposes, in contrast to the typically observational datasets in data mining, that were generated to support the operational setting and not for analysis in the first place. LMSs also employ a database and user authentication (password protection) which allows identifying the users in the logs. Some typical tasks of the data preparation phase are: data discretization (numerical values are transformed to categorical values), derivation of new attributes and selection of attributes (new attributes are created from the existed ones and only a subset of relevant attributes are chosen), creating summarization tables (these tables integrate all the desired information to be mined at an appropriate level, e.g. student), transforming the data format (to format required by the used data mining algorithms or frameworks).

- **Applying the mining algorithms.** In this step it is necessary: 1) to choose he specific association rule mining algorithm and implementation; 2) to configure the parameters of the algorithm, such as support and confidence threshold and others; 3) to identify which table or data file will be used for the mining; 4) and to specify some other restrictions, such as the maximum number of items and what specific attributes can be present in the antecedent or consequent of the discovered rules.

- **Data post-processing.** The obtained results or rules are interpreted, evaluated and used by the teacher for further actions. The final objective is to putting the results into use. Teachers use the discovered information (in form of if-then rules) for making decisions about the students and the LMS activities of the course in order to improve the students' learning. So, data mining algorithms have to express the output in a comprehensible format by e.g., using standardized e-learning metadata.

It is important to notice that traditional educational data sets are normally small [28] if we compare them to databases used in other data mining fields such as e-commerce applications that involve thousands of clients. This is due to the fact that the typical size of one classroom is often only between 10-100 students, depending on the type of the course (elementary, primary, adult, higher, tertiary, academic and special education). In the distance learning setting, the class size is usually larger, and it is also possible to pool data from several years or from several similar courses. Furthermore, the total number of instances or transactions can be quite large depending on how much information the LMS stores about the interaction of each student with the system (and at what levels of granularity). In this way, the number of available instances is much higher than the number of students. And, as we have said previously, educational data has also one advantage compared to several other domains [28]: the data sets are usually very clean, i.e., the values are correct and do not contain any noise from measuring devices.

3   Drawbacks and solutions

In the association rule mining area, most of the research efforts went in the first place to improving the algorithmic performance [29], and in the second place into reducing the output set by allowing the possibility to express constraints on the desired results. Over the past decade a variety of algorithms that address these issues through the refinement of search strategies, pruning techniques and data structures have been developed. While most algorithms focus on the explicit discovery of all rules that satisfy minimal support and confidence constraints 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 [30].

There are also other specific problems related to the application of association rule mining from e-learning data. When trying to solve these problems, one should consider the purpose of the association models and the data they come from. Nowadays, normally, data mining tools are designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex for educators to use and their features go well beyond the scope of what an educator might require. As a result, the courses administrator is more likely to apply data mining techniques in order to produce reports for instructors who then use these reports to make decisions about how to improve the student's learning and the online courses. However, it is most desirable that teachers participate directly in the iterative mining process in order to obtain more valuable rules. But normally, teachers only use the feedback provided by the obtained rules to make decisions about modification to improve the course, detect activities or students with problems, etc.

Some of the main drawbacks of association rule algorithms in e-learning are: the used algorithms have too many parameters for somebody non expert in data mining and the obtained rules are far too many, most of them non-interesting and with low comprehensibility. In the following subsections, we will tackle these problems.

3.1 Finding the appropriate parameter settings of the mining algorithm

Association rule mining algorithms need to be configured before to be executed. So, the user has to give appropriate values for the parameters in advance (often leading to too many or too few rules) in order to obtain a good number of rules. A comparative study between the main algorithms that are currently used to discover association rules can be found in [31]: Apriori [32], FP-Growth [33], MagnumOpus [34], Closet [35]. Most of these algorithms require the user to set two thresholds, the minimal support and the minimal confidence, and find all the rules that exceed the thresholds specified by the user. Therefore, the user must possess a certain amount of expertise in order to find the right settings for support and confidence to obtain the best rules.

One possible solution to this problem can be to use a parameter-free algorithm or with less parameters. For example, the Weka [36] package implements an Apriori-type algorithm that solves this problem partially. This algorithm reduces iteratively the minimum support, by a factor delta support (Δs) introduced by the user, until a minimum support is reached or a required number of rules (NR) has been generated.

Another improved version of the Apriori algorithm is the Predictive Apriori algorithm [37], which automatically resolves the problem of balance between these two parameters, maximizing the probability of making an accurate prediction for the data set. In order to achieve this, a parameter called the exact expected predictive accuracy is defined and calculated using the Bayesian method, which provides information about the accuracy of the rule found. In this way the user only has to specify the maximal number or rules to discover.

In [38] experimental tests were performed on a Moodle course by comparing the two previous algorithms. The final results demonstrated better performance for Predictive Apriori than Apriori-type algorithm using the Δs factor.

3.2 Discovering too many rules

The application of traditional association algorithms will be simple and efficient. However, association rule mining algorithms normally discover a huge quantity of rules and do not guarantee that all the rules found are relevant. Support and confidence factors can be used for obtaining interesting rules which have values for these factors grater than a threshold value. Although these two parameters allow the pruning of many associations, another common constraint is to indicate the attributes that must or cannot be present in the antecedent or consequent of the discovered rules.

Another solution is to evaluate, and post-prune the obtained rules in order to find the most interesting rules for a specific problem. Traditionally, the use of objective interestingness measures has been suggested [39], such as support and confidence, mentioned previously, as well as others measures such as Laplace, chi-square statistic, correlation coefficient, entropy gain, gini, interest, conviction, etc. These measures can be used for ranking the obtained rules in order than the user can select the rules with highest values in the measures that he/she is more interested.

Subjective measures are becoming increasingly important [40], in other words measures that are based on subjective factors controlled by the user. Most of the subjective approaches involve user participation in order to express, in accordance with his or her previous knowledge, which rules are of interest. Some suggested subjective measures [41] are:

- Unexpectedness: Rules are interesting if they are unknown to the user or contradict the user's knowledge.
- Actionability: Rules are interesting if users can do something with them to their advantage.

The number of rules can be decreased by only showing unexpected and actionable rules to the teacher and not all the discovered rules [38]. In [41], an Interestingness Analysis System (IAS) is proposed. It compares rules discovered with the user's knowledge about the area of interest. Let U be the set of user's specifications representing his/her knowledge space, A be the set of discovered association rules, this algorithm implements a pruning technique for removing redundant or insignificant rules by ranking and classifying them into four categories:

- Conforming rules: a discovered rule $A_i \in A$ conforms to a piece of user's knowledge U_j if both the antecedent and the consequent parts of A_i match those of $U_j \in U$ well.

- Unexpected consequent rules: a discovered rule $A_i \in A$ has unexpected consequents with respect to $U_j \in U$ if the antecedent part of A_i matches that of U_j well.
- Unexpected condition rules: a discovered rule $A_i \in A$ has unexpected conditions with respect to $U_j \in U$ if the consequent part of A_i matches that of U_j well, but not the antecedent part.
- Both-side unexpected rules: a discovered rule $A_i \in A$ is both-side unexpected with respect to $U_j \in U$ if the antecedent and consequent parts of A_i don't match those of U_j well.

The degrees of membership into each of these four categories are used for ranking the rules. Using their own specification language, they indicate their knowledge about the matter in question, through relationships among the fields or items in the database.

Finally, we can also use the knowledge database as a rule repository on the basis of which subjective analysis of the rules discovered is performed [38]. Before running the association rule mining algorithm, the teacher could download the relevant knowledge database, in accordance with his/her profile. The personalisation of the rules returned is based on filtering parameters, associated with the type of the course to be analysed such as: the area of knowledge; the level of education; the difficulty of the course, etc. The rules repository is created on the server in a collaborative way where the experts can vote for each rule in the repository, based on the educational considerations and their experience gained in other similar e-learning courses.

3.3 Discovery of poorly understandable rules

A factor that is of major importance in determining the quality of the extracted rules is their comprehensibility. Although the main motivation for rule extraction is to obtain a comprehensible description of the underlying model's hypothesis, this aspect of rule quality is often overlooked due to the subjective nature of comprehensibility, which can not be measured independently of the person using the system [42]. Prior experience and domain knowledge of this person play an important role in assessing the comprehensibility. This contrasts with accuracy that can be considered as a property of the rules and which can be evaluated independently of the users.

There are some traditional techniques that have been used in order to improve the comprehensibility of discovered rules. For example, we can reduce the size of the rules by constraining the number of items in the antecedent or consequent of the rule. Simplicity of the rule is related with its size, and as such, the shorter the rule is, the more comprehensible it will be. Another technique is performing a discretization of numerical values. Discretization [43] divides the numerical data into categorical classes that are easier to understand for the instructor (categorical values are more user-friendly for the instructor than precise magnitudes and ranges).

Another way to improve the comprehensibility of the rules is to incorporate domain knowledge and semantics, and to use a common and well-know vocabulary for the teacher. In the context of web-based educational systems, we can identify some common attributes to a variety of e-learning systems such as LMS and adaptive hypermedia courses. As we can see in table 1, these attributes could be present in many sections or levels of the course. For example, a unit could be a chapter, or a lesson, or an exercise, or a collaborative resource.

Table 1. Examples of attributes common to a variety of e-learning systems.

Attribute	Description
Visited	If the unit, document or web page has been visited
Total_time	Time taken by the student to complete the unit
Score	Average final score for the unit
Knowledge_level	Student's initial and final level in the unit
Difficulty_level	Difficulty level of the unit
Attempts	Number of attempts before passing the unit
Chat_messages	Number of messages sent/read in the chat room
Forum_messages	Number of messages sent/read in the forum

In this context the use of standard metadata for e-learning [44] allows the creation and maintenance of a common knowledge base with a common vocabulary susceptible of sharing among different communities of instructors. For example, the SCORM [45] standard describes a content aggregation model and a tracking model for reusable learning objects. Although SCORM provides a framework for the representation and processing of the metadata, it falls short in including the support needed for other, more specific, pedagogical tracking such as the use of collaborative resources. In Figure 1, we show a proposed SCORM-based Ontology for association rule mining in LMS using standard e-learning attributes.
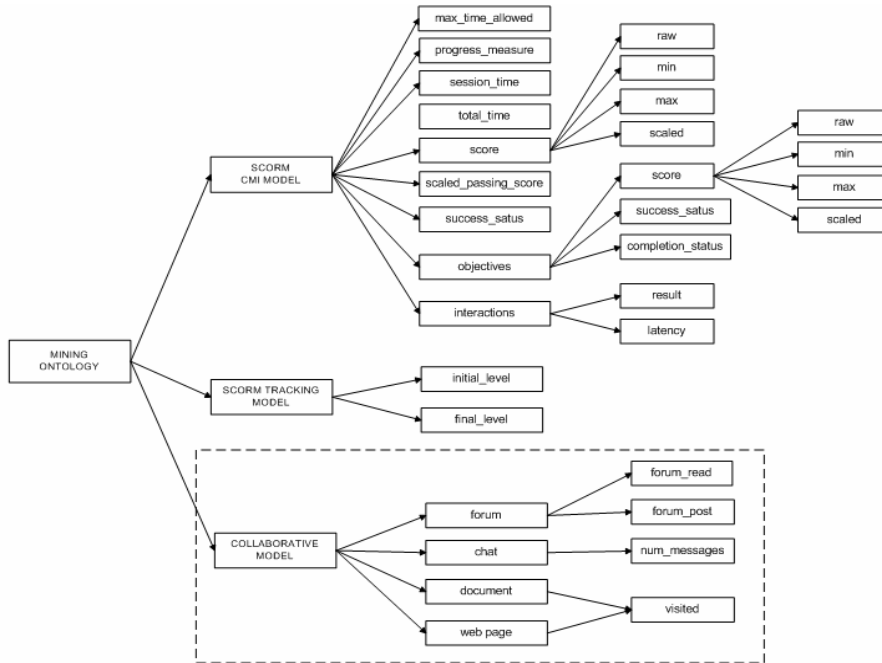
Fig. 1. SCORM based Ontology for association rule mining in LMS.

The ontology of Figure 1 includes, besides the standard SCORM attributes, other attributes related to the collaborative learning. This could be a good starting point for content re-using and for exchanging results between different mining frameworks.

In order to improve the comprehensibility and suitability of the rules, it will be very useful to also provide an ontology that describes the specific domain [44]. In this way the teacher can understand better the rules that contain concepts related to the domain under study, like "*if success in topic A then success in topic B.*"

Finally, another proposal is to use domain specific interactive data mining [46] in which the user is involved in the discovery process to find iteratively the most interesting results. Domain and problem specific representation are also added to the mining process. The user is not just evaluating the result of an automatic data mining process, but he or she is actively involved in the design of a new representation and the search for patterns.

4 Conclusions and future trends

It is still the early days for the total integration of association rule mining in e-learning systems and not many real and fully operative implementations are available. In this paper, we have outlined some of the main drawbacks for the application of association rule mining in learning management systems and we have described some possible solutions for each problem.

We believe that some future research lines will focus on: developing association rule mining tools that can more easily be used by educators; proposing new specific measures of interest with the inclusion of domain knowledge and semantic; embedding and integrating mining tools into LMSs in order to enable the teacher to use the same interface to create/maintain courses and to carry out the mining process/obtain direct feedback/make modifications in the course; developing iterative and interactive or guided mining to help educators to apply KDD processes, or even developing an automatic mining system that can perform the mining automatically in an unattended way, such that the teacher only has to use the proposed recommendations in order to improve the students' learning.

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).

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.
9. 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.
10. Merceron, A., & Yacef, K.: Mining student data captured from a web-based tutoring tool. Journal of Interactive Learning Research, 15:4 (2004) 319–346.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. Ha, S., Bae, S., Park, S.: Web Mining for Distance Education. In: IEEE Int. Conf. on Management of Innovation and Technology (2000) 715–719.
17. Wang, W., Weng, J., Su, J., Tseng, S.: Learning portfolio analysis and mining in scorm compliant environment. In: ASEE/IEEE Frontiers in Education Conf. (2004) 17–24.
18. Machado, L., Becker, K.: Distance Education: A Web Usage Mining Case Study for the Evaluation of Learning Sites. In: Proc. of Int. Conf. on Advanced Learning Technologies (2003) 360-361.
19. Kay, J., Maisonneuve, N., Yacef, K., Zaiane, O.R. : Mining Patterns of Events in Students' Teamwork Data. In: Proc. of Educational Data Mining Workshop (2006) 1-8.
20. 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.
21. Tsai, C.J., Tseng, S.S., Lin, C.Y.: A Two-Phase Fuzzy Mining and Learning Algorithm for Adaptive Learning Environment. In: Proc. of Int. Conf. on Computational Science (2001) 429-438.
22. Hwang, G.J., Hsiao, C.L., Tseng, C.R.: A Computer-Assisted Approach to Diagnosing Student Learning Problems in Science Courses. Journal of Information Science and Engineering 19 (2003) 229-248.
23. Kumar, A.: Rule-Based Adaptive Problem Generation in Programming Tutors and its Evaluation. In: Proc. of Int. Conf. on Artificial Intelligence in Education (2005) 36-44.
24. Matsui, T., Okamoto, T.: Knowledge Discovery from Learning History Data and its Effective Use for Learning Process Assessment Under the e-Learning Environment. In:

Proc. of Int. Conf. on Society for Information Technology and Teacher Education (2003) 3141-3144.
25. Costabile, M.F., De Angeli, A., Roselli, T., Lanzilotti, R., Plantamura, P.: Evaluating the Educational Impact of a Tutoring Hypermedia for Children. In: Proc. of Int. Conf. Information Technology in Childhood Education Annual (2003) 289-308.
26. Hsu, H.H., Chen, C.H., Tai, W.P.: Towards Error-Free and Personalized Web-Based Courses. In: Proc. of Int. Conf. on Advanced Information Networking and Applications (2003) 99-104.
27. Dos Santos, M.L., Becker, K.: Distance Education: a Web Usage Mining Case Study for the Evaluation of Learning Sites. In: IEEE Int. Conf. on Advanced Learning Technologies (2003) 360-361.
28. 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.
29. 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.
31. Zheng Z., Kohavi R., Mason L. Real world performance of association rules. In: Proc. of the Sixth ACM-SIGKDD (2001) 86-98.
32. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc. of Int. Conf. on Very Large Data Bases (1996) 487-499.
33. 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.
34. Webb, G.I.: OPUS: An efficient admissible algorithm for unordered search. Journal of Artificial Intelligence Research 3 (1995) 431-465.
35. 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.
36. Weka. (2007) Available at http://www.cs.waikato.ac.nz/ml/weka/.
37. Tobias S.: Finding Association Rules That Trade Support Optimally against Confidence. In: Proc. of the European Conf. of PKDD (2001) 424-437.
38. García, E., Romero, C., Ventura, S. Castro, C.: Using Rules Discovery for the Continuous Improvement of e-Learning Courses. In Proc. of the Int. Conf. on Intelligent Data Engineering and Automated Learning (2006) 887-895.
39. Tan P., Kumar V.: Interesting Measures for Association Patterns: A Perspective. TR00-036. Department of Computer Science. University of Minnnesota. (2000) 1-36.
40. Silberschatz, A., Tuzhilin, A.: What makes pattterns interesting in Knoledge discovery systems. IEEE Trans. on Knowledge and Data Engineering, 8:6 (1996) 970-974.
41. Liu B., Wynne H., Shu C. Yiming M.: Analyzing the Subjective Interestingness of Association Rules. IEEE Intelligent System 15:5 (2000) 47-55.
42. Huysmans J., Baesens B., Vanthienen J.: Using Rule Extraction to Improve the Comprehensibility of Predictive Models. FETEW Research Report (2006) 1-55.
43. Dougherty, J. Kohavi, M. Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Int. Conf. Machine Learning (1995) 194–202.
44. Brase J., Nejdl W.: Ontologies and Metadata for eLearning. In S. Staab & R. Studer (Eds.) Handbook on Ontologies, Springer Verlag, (2003) 579-598.
45. ADL: Advanced Distributed Learning. Shareable content object reference model (SCORM): The SCORM overview. Available at http://www.adlnet.org
46. Hubscher, R., Puntambekar, S., Nye, A.: Domain Specific Interactive Data Mining. Workshop on Data Mining for User Modeling at UM'07 (2007) 81-90.