Educational Data Mining: A Review of the State-of-the-Art

Cristóbal Romero, *Member, IEEE*, Sebastián Ventura, *Senior Member, IEEE*

***Abstract*—Educational Data Mining is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. This paper surveys the most relevant studies carried out in this field to date. Firstly, it introduces EDM and describes the different groups of user, types of educational environments and the data they provide. It then goes on to list the most typical/common tasks in the educational environment that have been resolved through data mining techniques and finally some of the most promising future lines of research are discussed.**

***Index Terms*— Educational Data Mining, Knowledge Discovery, Educational Systems, Data Mining.**

I. Introduction

EDUCATIONAL Data Mining (EDM) is the application of Data Mining (DM) techniques to educational data, and so, its objective is to analyze these type of data in order to resolve educational research issues [27].

DM can be defined as the process involved in extracting interesting, interpretable, useful and novel information from data [78]. It has been used for many years by businesses, scientists and governments to sift through volumes of data like airline passenger records, census data and the supermarket scanner data that produces market research reports [105].

EDM is concerned with developing methods to explore the unique types of data in educational settings and, using these methods, to better understand students and the settings in which they learn [21]. On one hand, the increase in both instrumental educational software as well as state databases of student information has created large repositories of data reflecting how students learn [145]. On the other hand, the use of Internet in education has created a new context known as e-learning or web-based education in which large amounts of information about teaching-learning interaction are endlessly generated and ubiquitously available [60]. All this information provides a gold mine of educational data [188]. EDM seeks to use these data repositories to better understand learners and learning, and to develop computational approaches that combine data and theory to transform practice to benefit learners. EDM has emerged as a research area in recent years for researchers all over the world from different and related research areas such as:

- Offline education try to transmit knowledge and skills based on face-to-face contact and also study psychologically on how humans learn. Psychometrics and statistical techniques have been applied to data like student behavior/performance, curriculum, etc. that was gathered in classroom environments
- E-learning and Learning Management System (LMS). E-learning provides online instruction and LMS also provides communication, collaboration, administration and reporting tools. Web Mining (WM) techniques have been applied to student data stored by these systems in log files and databases.
- Intelligent Tutoring (ITS) and Adaptive Educational Hypermedia System (AEHS) are an alternative to the just-put-it-on-the-web approach by trying to adapt teaching to the needs of each particular student. Data Mining has been applied to data picked up by these systems, such as log files, user models, etc.

The EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. This process does not differ much from other application areas of data mining like business, genetics, medicine, etc. because it follows the same steps as the general data mining process [221]: pre-processing, data mining and post-processing. However, it is important to notice that in this paper the term data mining is used in a larger sense than the original/traditional DM definition. That is, we are going to describe not only EDM studies that use typical DM techniques such as classification, clustering, association rule mining, sequential mining, text mining, etc. but also other approaches such as regression, correlation, visualization, etc. that are not considered to be DM in a strict sense. Furthermore, some methodological innovations and trends in EDM such as discovery with models and the integration of psychometric modeling frameworks are unusual DM categories or not necessarily universally seen as being DM [20].

From a practical view point EDM allows, for example, to discover new knowledge based on students' usage data in order to help validate/evaluate educational systems, to

Cristobal Romero is with the Cordoba University, Campus de Rabanales, 14071 Córdoba, Spain (phone: 34-957-212172; fax: 34-957-218630; e-mail: comero@ uco.es).
Sebastián Ventura is with the Cordoba University, Campus de Rabanales, 14071 Córdoba, Spain (e-mail: sventura@ uco.es).

potentially improve some aspects of the quality of education and to lay the groundwork for a more effective learning process [221]. Some similar ideas were already successfully applied in e-commerce systems, the first and most popular application of data mining [213], in order to determine clients' interests so as to be able to increase online sales. However, there has been comparatively less progress in this direction in Education to date, although this situation is changing and there is currently an increasing interest in applying data mining to the educational environment [230]. Even so, there are some important issues that differentiate the application of DM specifically to education from how it is applied in other domains [223]:

- Objective. The objective of data mining in each application area is different. For example, in business the main objective is to increase profit, which is tangible and can be measured in term of amounts of money, number of customers and customer loyalty. But EDM has both applied research objectives, such as improving the learning process and guiding students' learning; as well as pure research objectives, such as achieving a deeper understanding of educational phenomena. These goals are sometimes difficult to quantify and require their own special set of measurement techniques.

- Data. In educational environments there are many different types of data available for mining. These data are specific to the educational area and so have intrinsic semantic information, relationships with other data and multiple levels of meaningful hierarchy. Some examples are the domain model, used in ITS and AEHS, that represents the relationships among the concepts of a specific subject in a graph or hierarchy format (e.g. a course consists of several chapters that are organized in lessons and each lesson includes several concepts); and the q-matrix that shows relationships between items/questions of a test/quiz system and the concepts evaluated by the test. Furthermore, it is also necessary to take pedagogical aspects of the learner and the system into account.

- Techniques. Educational data and problems have some special characteristics that require the issue of mining to be treated in a different way. Although most of the traditional DM techniques can be applied directly, others cannot and have to be adapted to the specific educational problem at hand. Furthermore, specific data mining techniques can be used for specific educational problems.

EDM involves different groups of users or participants. Different groups look at educational information from different angles according to their own mission, vision and objectives for using data mining [106]. For example, knowledge discovered by EDM algorithms can be used not only to help teachers to manage their classes, understand their students' learning processes and reflect on their own teaching methods, but also to support a learner's reflections on the situation and provide feedback to learners [179]. Although an initial consideration seems to involve only two main groups, the learners and the instructors, there are actually more groups involved with many more objectives, as can be seen in Table I.

TABLE I
EDM USERS/STAKEHOLDERS.

Users/Actors	Objetives for using data mining
Learners/ Students/ Pupils	To personalize e-learning; to recommend activities to learners and resources and learning tasks that could further improve their learning; to suggest interesting learning experiences to the students; to suggest path pruning and shortening or simply links to follow, to generate adaptive hints, to recommend courses, relevant discussions, books, etc.
Educators/ Teachers/ Instructors/ Tutors	To get objective feedback about instruction; to analyze students' learning and behavior; to detect which students require support; to predict student performance; to classify learners into groups; to find a learner's regular as well as irregular patterns; to find the most frequently made mistakes; to determine more effective activities; to improve the adaptation and customization of courses, etc.
Course Developers/ Educational Researchers	To evaluate and maintain courseware; to improve student learning; to evaluate the structure of course content and its effectiveness in the learning process; to automatically construct student models and tutor models; to compare data mining techniques in order to be able to recommend the most useful one for each task; to develop specific data mining tools for educational purposes; etc.
Organizations/ Learning Providers/ Universities/ Private Training Companies	To enhance the decision processes in higher learning institutions; to streamline efficiency in the decision-making process; to achieve specific objectives; to suggest certain courses that might be valuable for each class of learners; to find the most cost-effective way of improving retention and grades; to select the most qualified applicants for graduation; to help to admit students who will do well in university, etc.
Administrators/ School District Administrators/ Network Administrators/ System Administrators	To develop the best way to organize institutional resources (human and material) and their educational offer; to utilize available resources more effectively; to enhance educational program offers and determine the effectiveness of the distance learning approach; to evaluate teacher and curricula; to set parameters for improving web-site efficiency and adapting it to users (optimal server size, network traffic distribution, etc.).

Nowadays, there is a great variety of educational systems/environments such as: the traditional classroom, e-learning, LMS, AH educational systems, ITS, tests/quizzes, texts/contents, and others such as: learning object repositories, concept maps, social networks, forums, educational game environments, virtual environments, ubiquitous computing environments, etc. All data provided by each of the above-mentioned educational environments are different, thus enabling different problems and tasks to be resolved using data mining techniques (see Section II). Table II shows a list of the most important studies on EDM grouped according to the type of data/environment involved.

On the other hand, the International Working Group in EDM (*http://www.educationaldatamining.org*) has achieved the establishment of an annual International Conference on Educational Data Mining in 2008, EDM08 [19], EDM09 [27], EDM10 [22]. This conference has evolved from previous EDM Workshops at the AIED07 [114], EC-TEL07 [224], ICALT07 [35], UM07 [17], AAII06 [34], ITS06 [113], AAAI05 [33], AIED05 [62], ITS04 [32] and ITS00 [30] conferences.

TABLE II
LIST OF EDM REFERENCES GROUPED ACCORDING TO TYPES OF DATA USED.

Type of Data/ Environment	References
Traditional Education	[32], [42], [66], [68], [79], [95], [98], [103], [119], [120], [123], [130], [133], [141], [142], [147], [148], [164], [165], [169], [175], [197], [198], [212], [217], [238], [239], [241], [254], [260], [263], [271], [273], [280], [292], [306].
Web-based Education/ E-learning	[11], [45], [49], [50], [63], [64], [86], [92], [97], [100], [102], [104], [118], [122], [129], [132], [146], [149], [153], [155], [156], [157], [158], [159], [177], [181], [182], [183], [190], [193], [199], [201], [214], [216], [227], [240], [242], [248], [255], [261], [265], [274], [277], [278], [286], [287], [288], [290], [291], [294], [295], [297], [300], [302].
Learning Management Systems	[28], [46], [48], [59], [67], [76], [101], [111], [112], [134], [161], [166], [170], [173], [180], [184], [185], [210], [211], [225], [226], [234], [244], [256], [268], [269], [276], [293], [305].
Intelligent Tutoring Systems	[9], [15], [16], [18], [26], [29], [31], [47], [61], [65], [84], [99], [108], [116], [126], [136], [145], [176], [179], [187], [202], [205], [215], [219], [220], [236], [251], [267], [282], [289], [296].
Adaptive Educational Systems	[4], [23], [37], [38], [69], [93], [94], [107], [125], [127], [135], [138], [140], [150], [162], [163], [189], [221], [229], [247], [259], [262], [270], [279], [281], [303].
Tests/ Questionnaires/	[7], [12], [14], [25], [41], [43], [51], [54], [57], [80], [89], [128], [167], [196], [203], [204], [206], [207], [250], [272], [283], [285], [304].
Texts/ Contents	[1], [3], [40], [73], [109], [143], [152], [160], [237], [249], [253], [266], [285], [299].
Others	[2], [13], [44], [53], [55], [71], [74], [77], [110], [124], [139], [144], [154], [192], [200], [208], [218], [233], [235], [252], [264], [301].

The number of publications about EDM has grown exponentially in the last few years (see Figure 1). A clear sign of this tendency is the appearance of the peer-reviewed journal JEDM (Journal of Educational Data Mining) and two specific books on EDM edited by Romero & Ventura entitled: Data Mining in E-learning [222] and The Handbook of Educational Data Mining [230] co-edited with Baker & Pechenizkiy. There were also two surveys carried out previously about EDM. The first one [223] is a former review of Romero & Ventura with 81 references until 2005 in which papers were classified by the DM techniques used. In fact, the present survey is an improved, updated and much extended version of this previous one with 306 references in which papers are classified by educational categories/tasks and the types of data used. It also shows some examples of new categories that have appeared since the 2005 survey such as social network analysis and constructing courseware. The other survey [20] is a recent review by Ryan & Yacef with 46 references encompassing up to 2009. This survey uses mainly the top 8 most cited papers in the first 2005 review and the proceedings of EDM08 and EDM09 conferences; it also groups papers according to EDM methods and applications, as we describe in the next section.

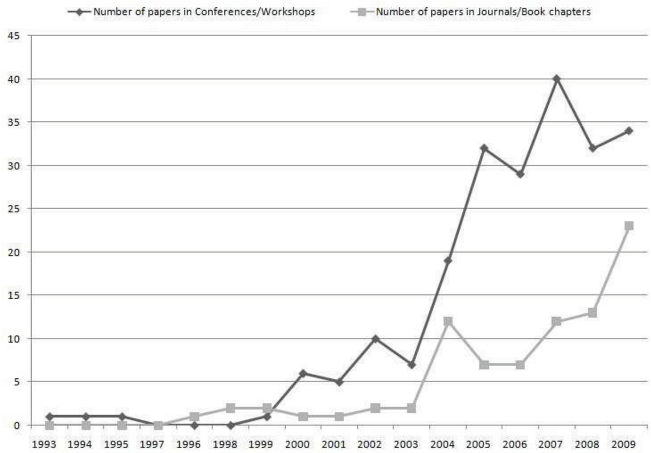

Fig. 1. Number of published papers until 2009 grouped according to the year. Notice that we have counted only the three hundred papers in our reference section and not the total number of papers that were really published about EDM.

This survey is organized as follows: Section II lists the most common tasks in education that have been resolved by using data mining techniques. Section III, describes some of the most prominent future research lines. Finally, conclusions are outlined in Section IV.

II. EDUCATIONAL TASKS AND DATA MINING TECHNIQUES

There are many applications or tasks in educational environments that have been resolved through DM. For example, Baker [20],[21] suggests four key areas of application for EDM: improving student models, improving domain models, studying the pedagogical support provided by learning software, scientific research into learning and learners; and five approaches/methods: prediction, clustering, relationship mining, distillation of data for human judgment and discovery with models. Castro [60] suggests the following EDM subjects/tasks: applications dealing with the assessment of the student's learning performance, applications that provide course adaptation and learning recommendations based on the student's learning behavior, approaches dealing with the evaluation of learning material and educational web-based courses, applications that involve feedback to both teacher and students in e-learning courses, and developments for detection of atypical students' learning behaviors. However, as we think that there are even more possible applications, we have established our own categories (see Figure 2) for the main educational tasks which have employed data mining techniques. These categories come from different research communities (as we have previously described in the Introduction) and they also use different DM tasks and techniques. On the one hand, we can see in Table II that the most active communities are e-learning/LMS and ITS/AEHS. On the other hand, we will see in the following subsections that the most commonly applied DM tasks are regression, clustering, classification and association rule mining; and the most used DM techniques/methods are decision trees, neural networks and bayesian networks.

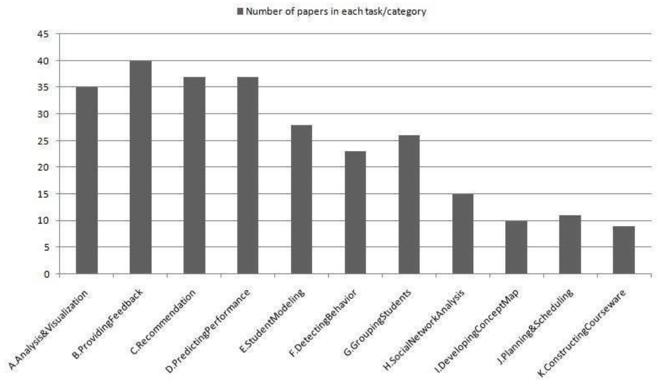

Fig. 2. Number of published papers until 2009 grouped by task/category. Notice that we have counted only the three hundred papers inour reference section and not the total number of papers actually published about EDM.

As we can see in Figure 2, the categories or research lines that have had the most papers published are the first 8 ones (from A to G with 23 or more references each) and the categories that have the fewest papers published are the last 4 (from H to K with less than 15 references). We think that this may be due mainly to the fact that the first 8categories are older than the last 4 (and so more authors have worked on these tasks) but it could also be because of the special interest in each one. For example, although social network analysis is one of the newest tasks, it has more papers than the other 3. We also want to point out, that we have organised these categories by grouping them near the most closely related ones, that in our opinion are the following since tasks A and B provide information to instructors and C to the students; D, E, F and G tasks reveal students' characteristics; H and I study graphs and relationships between students and concepts respectively; and J and K help in creating/planning courseware and the course, respectively. Next, we are going to describe in detail these tasks/categories and the most relevant studies. But, as there are closely related areas, some references could be located in a different category or in several.

A. Analysis and visualization of data

The objective of the analysis and visualization of data is to highlight useful information and support decision making. In the educational environment, for example, it can help educators and course administrators to analyze the students' course activities and usage information to get a general view of a student's learning. Statistics and visualization information are the two main techniques that have been most widely used for this task.

Statistics is a mathematical science concerning the collection, analysis, interpretation or explanation, and presentation of data [87]. It is relatively easy to get basic descriptive statistics from statistical software such as SPSS. Used with educational data, this descriptive analysis can provide such global data characteristics as summaries and reports about learner behavior [284]. It is not surprising that teachers prefer pedagogically oriented statistics (overall success rate, mastery levels, typical misconceptions, percentage of exercises tackled and material read) which are easy to interpret [303]. On the other hand, teachers find the fine-grained statistics in log data too cumbersome to inspect or too time-consuming to interpret. Statistical analysis of educational data (logs files/databases) can tell us such things as: where students enter and exit, the most popular pages, the browsers students tend to use, patterns of use over time, [132]; the number of visits, origin of visitors, number of hits, patterns of use throughout various time periods [96]; number of visits and duration per quarter, top search terms, number of downloads of e-learning resources [100]; number of different pages browsed, total time for browsing the different pages [129]; usage summaries and reports on weekly and monthly user trends and activities [185]; session statistics and session patterns [201]; statistical indicators on the learner's interactions in forums [5]; the amount of material students might go through, the order in which students study topics [214]; resources used by students, resources valued by students [243]; the overall averages of contributions to discussion forums, the amount of posting vs. replies, the amount of learner-to-learner interaction vs. learner-to-teacher interaction [112]; the time a student dedicates to the course or a particular part of it [201]; the learners' behavior and time distribution , the distribution of network traffic over time [305]; the frequency of studying events, patterns of studying activity, timing and sequencing of events and the content analysis of students' notes and summaries [104]. Statistical analysis is also very useful to obtain reports assessing [82] how many minutes the student has worked, how many minutes he has worked today, how many problems he has resolved and his correct percentage, our prediction of his score and his performance level.

Information visualization uses graphic techniques to help people understand and analyze data [174]. Visual representations and interaction techniques take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once. There are several studies oriented toward visualizing different educational data such as: patterns of annual, seasonal, daily and hourly user behavior on online forums [40]; the complete educational (assessment) process [207]; mean values of attributes analyzed in data to measure mathematical skills [304]; tutor-student interaction data from an automated reading tutor [187]; statistical graphs about assignments complement, questions admitted, exam score and so on [244]; student tracking data regarding social, cognitive and behavioral aspects of students [172]; student attendance, access to resources, overview of discussions and results on assignments and quizzes [173]; weekly information regarding students' and groups' activity [137]; student progression per question as a transition between the types of questions [38]; fingertip actions in collaborative learning activities [11]; deficiencies in a student's basic understanding of individual concepts [288] and higher- education student-evaluation data [133]; student's interactions with online learning environments [134]; the students' on-line exercise work including students' interactions and answers, mistakes, teachers' comments and so on [178]; questions and suggestions in an adaptive tutorial [39]; navigational behavior and the performance of the learner [37]; educational trails of Web-pages visited and activities done [227] and the sequence of learning objects and educational trails [240].

B. Providing feedback for supporting instructors

The objective is to provide feedback to support course authors/teachers/administrators in decision making (about how to improve students' learning, organize instructional resources more efficiently, etc) and enable them to take appropriate proactive and/or remedial action. It is important to point out that this task is different than data analyzing and visualizing tasks, which only provide basic information directly from data (reports, statistics, etc.). Moreover, providing feedback divulges completely new, hidden and interesting information found in data. Several DM techniques have been used in this task, although association rule mining has been the most common. Association rule mining reveals interesting relationships among variables in large databases and presents them in the form of strong rules according to the different degrees of interest they might present [298].

There are many studies that apply/compare several data mining models that provide feedback. Association rules, clustering, classification, sequential pattern analysis, dependency modeling and prediction have been used to enhance web-based learning environments to improve the degree to which the educator can evaluate the learning process [294]. Association analysis, clustering analysis and case-based reasoning have also been used to organize course material and assign homework at different levels of difficulty [245]. Clustering, classification and association rule mining have been applied to develop a service to allow the evaluator to gather feedback from the learning progress automatically and thus appraise online course effectiveness [234]. Decision trees, Bayesian models and other prediction techniques have been proposed to address the admission and counseling process in order to assist in improving the quality of education and student performance [217]. Several classifier algorithms have been applied to predict whether the teacher will recommend an intervention strategy for motivational profiles [126]. Clustering and association rules have been used in the academic community to potentially improve some qualitative teaching aspects [273].

Association rule mining has been used to confront the problem of continuous feedback in the educational process [210]; to analyze learning data and to figure out whether students use resources and possibly whether their use has any (positive) impact on marks [180]; to determine the relationship between each learning-behavior pattern so that the teacher can promote collaborative learning behavior on the Web [291]; to find embedded information, which can be provided to teachers to further analyze, refine or reorganize teaching materials and tests in adaptive learning environments [262]; to optimize the content of the university e-learning portal [216]; to discover interesting associations between student attributes, problem attributes and solution strategies in order to improve online education systems for both teachers and students [183]; to analyze rule evaluation measures in order to discover the most interesting rules [269]; to identify interesting and unexpected learning patterns which in turn may provide decision lines enabling teachers to more efficiently organize their teaching structure [274]; to provide feedback to the course author about how to improve courseware [221]; to analyze the user's access log in Moodle to improve e-e-learning and to support the analysis of trends [28]; to find relationships between students' LMS access behavior and overall performances in order to understand student web usage patterns [46]; to improve an adaptive course design in order to show recommendations on how to enhance the course structure and contents [270]; to find interesting relationships between attributes, solution strategies adopted by learners and so on, from a web-based mobile learning system [301]; to help the teacher to discover beneficial or detrimental relationships between the use of web-based educational resources and student learning [228]; to reveal information about university student enrollment [238]; to help organizations determine the thinking styles of learners and the effectiveness of a web site structure [102]; to evaluate educational web site design [166] and to mine open answers in questionnaire data in order to analyze surveys [285].

Other different DM techniques have been applied to provide feedback, such as: domain specific interactive data mining to find the relationships between log data and student behavior in an educational hypermedia system [125]; temporal data mining to describe, interpret and predict student behavior, and to evaluate progress in relation to learning outcomes in ITSs [29]; learning decomposition and logistic regression to compare the impact of different educational interventions on learning [85]; timely alerts to detect critical teaching and learning patterns and to help teachers make sense of what is happening in their classrooms [248]; usage data analysis to improve the effectiveness of the learning process in e-learning systems [184].

A special type of feedback is when data come specifically from tests, questions, assessments, etc. In this case the objective is to analyze it in order to improve the questionnaires and to answer questions such as: what items/questions test the same information and which are of the most use for predicting course/test results etc. Several DM approaches and techniques (clustering, classification and association analysis) have been proposed for joint use in the mining of student assessment data [206]. A group of data mining techniques, i.e. statistic correlation analysis, fuzzy clustering analysis, grey relational analysis, k-means clustering and fuzzy association rule mining have been applied to support mobile formative assessment in order to help teachers understand the main factors influencing learner performance [55]. Several clustering algorithms (k-means, agglomerative clustering and spectral clustering) have been applied to extract underlying relationships from a score matrix in order to help instructors to generate a large unit test [250]. Hierarchical clustering has been used for mining multiple-choice assessment data for similarity of the concepts represented by the responses [167]. Common-factor analysis and collaborative filtering have been used to discover the fundamental topics of a course from item-level grades [283]. Association rule mining has been applied to analyze questionnaire data by discovering rule patterns in questionnaire data [54].

Finally, another special type of feedback involves the use of text data. In this case, the objective of applying text/data mining to educational data is to analyze educational contents, to summarize/analyze the learner discussion process, etc. in order to provide instructor feedback. Automatic text analysis, content analysis and text mining have been used to extract and identify the opinions found on web pages in e-learning systems [249]; to mine free-form spoken responses given to tutor prompts by estimating the probability that a response has of mentioning a given target or set of targets [299]; to facilitate the automatic coding process of an online discussion forum [160]; for collaborative learning prompted by learners' comments on discussion boards [266]; to assess asynchronous discussion forums in order to evaluate the progress of a thread discussion [73]; and to identify patterns of interaction and their sequential organization in computer-supported collaborative environments like chats [44].

C. Recommendations for students

The objective is to be able to make recommendations directly to the students with respect to their personalized activities, links to visits, the next task or problem to be done, etc. and also to be able to adapt learning contents, interfaces and sequences to each particular student. Several DM techniques have been used for this task but the most common are association rule mining, clustering and sequential pattern mining. Sequence/Sequential pattern mining aims to discover the relationships between occurrences of sequential events, to find if there exists any specific order in the occurrences [70].

Sequential pattern mining has been developed to personalize recommendations on learning content based on learning style and web usage habits [300]; to study eye movements (of students reading concept maps) in order to detect when focal actions overlap unrelated actions [194]; for developing personalized learning scenarios in which the learners are assisted by the system based on patterns and preferred learning styles [23]; to identify significant sequences of activity indicative of problems/success in order to assist student teams by early recognition of problems [139]; to generate personalized activities for learners [279]; for personalizing based on itineraries and long-term navigational behavior [186]; to recommend the most appropriate future links for a student to visit in a web-based adaptive educational system [229]; to include the concept of recommended itinerary in SCORM standard by combining teachers' expertise with learned experience [186]; to select different learning objects for different learners based on learner profiles and the internal relation of concepts [246]; for personalizing activity trees according to learning portfolios in a SCORM compliant environment [279]; for recommending lessons (learning objects or concepts) that a student should study next while using an adaptive hypermedia system [150]; to discover LO relationship patterns to recommend related learning objects to learners [200]; for adapting learning resource sequencing [138].

Association rule mining has been used to recommend on-line learning activities or shortcuts on a course web site [295]; to produce recommendations for learning material in e-learning systems [168]; for content recommendation based on educationally-contextualized browsing events for web-based personalized learning [276]; for recommending relevant discussions to the students [2]; to provide students with personalized learning suggestions by analyzing their test results and test related concepts [57]; for making recommendations to courseware authors about how to improve adaptive courses [93]; for building a personalized e-learning material-recommender system to help students find learning materials [162]; for course recommendation with respect to optimal elective courses [255]; for designing a material recommendation system based on the learning actions of previous learners [161].

Clustering has been developed to establish a recommendation model for students in similar situations in the future [278]; for grouping web documents using clustering methods in order to personalize e-learning based on maximal frequent item sets [253]; for providing personalized course material recommendations based on learner ability [163] and to recommend to students those resources they have not yet visited but would find most helpful [97].

Other DM techniques used are: neural networks and decision trees to provide adaptive and personalized learning support [101]; production rules to help students to make decisions about their academic itineraries [271]; decision tree analysis to recommend optimal learning sequences to facilitate the students' learning process and maximize their learning outcome [281]; learning factor transfers and Q-matrixes to generate domain models that will sequence item-types to maximize learning [205]; an item-order effect model to suggest the most effective item sequences to facilitate learning [204]; a fuzzy item-response theory to recommend appropriate courseware for learners [50]; intelligent agent technology and SCORM based course objects to build an agent-based recommender system for lesson plan sequencing in web-based learning [286]; data mining and text mining to recommend books related to the books that the target pupil has consulted [191]; case-based reasoning to offer contextual help to learners, providing them with an adapted link structure for the course [116]; Markov decision process to automatically generate adaptive hints in ITS (to identify the action that will lead to the next state with the highest value) [251] and an extended Serial Blog Article Composition Particle Swarm Optimization (SBACPSO) algorithm to provide optimal recommended materials to users in blog-assisted learning [124].

D. *Predicting student performance*

The objective of prediction is to estimate the unknown value of a variable that describes the student. In education the values normally predicted are performance, knowledge, score or mark. This value can be numerical/continuous value (regression task) or categorical/discrete value (classification task). Regression analysis finds the relationship between a dependent variable and one or more independent variables [72]. Classification is a procedure in which individual items are placed into groups based on quantitative information regarding one or more characteristics inherent in the items and based on a training set of previously labeled items [75]. Prediction of a student's performance is one of the oldest and most popular applications of DM in education, and different techniques and models have been applied (neural networks, Bayesian networks, rule-based systems, regression and correlation analysis).

A comparison of machine learning methods has been carried out to predict success in a course (either passed or failed) in Intelligent Tutoring Systems [108]. Other comparisons of different data mining algorithms are made to classify students (predict final marks) based on Moodle usage data [226]; to predict student performance (final grade) based on features extracted from logged data [182] and to predict University students' academic performance [130].

Different types of neural network models have been used to predict final student grades (using back-propagation and feed-forward neural networks) [95]; to predict the number of errors a student will make (using feed-forward and backpropagation) [282]; to predict performance from test scores (using back-propagation and counter-propagation) [79]; to predict students' marks (pass or fail) from Moodle logs (using radial basis functions) [67] and for predicting the likely performance of a candidate being considered for admission into the university (using multilayer perceptron topology) [198].

Bayesian networks have been used to predict student-applicant performance [103]; to model user knowledge and predict student performance within a tutoring system [202]; to predict a future graduate's cumulative Grade Point Average based on applicant background at the time of admission [119]; to model two different approaches to determine the probability a multi skill question has of being corrected [203] and to predict future group performance in face-to-face collaborative learning [252]; to predict end-of-year exam performance through student activity with online tutors [12] and to predict item response outcome [69].

Different types of rule-based systems have been applied to predict student performance (mark prediction) in an e-learning environment (using fuzzy association rules) [193]; to predict learner performance based on the learning portfolios compiled (using key formative assessment rules) [51]; for prediction, monitoring and evaluation of student academic performance (using rule induction) [197]; to predict final grades based on features extracted from logged data in an education web-based system (using genetic algorithm to find association rules) [242]; to predict student grades in LMSs (using grammar guided genetic programming) [293]; to predict student performance and provide timely lessons in web-based e-learning systems (using decision tree) [45]; to predict online students' marks (using an orthogonal search-based rule extraction algorithm) [76].

Several regression techniques have been used to predict students' marks in an open university (using model trees, neural networks, linear regression, locally weighed linear regression and support vector machines) [148]; for predicting end-of-year accountability assessment scores (using linear regression prediction models) [7]; to predict student performance from log and test scores in web-based instruction (using a multivariable regression model) [290]; for predicting student academic performance (using stepwise linear regression) [98]; for predicting time to be spent on a learning page (using multiple linear regression) [8]; for identifying variables that could predict success in colleges courses (using multiple regression) [169]; for predicting university students' satisfaction (using regression and decision trees analysis) [260]; for predicting exam results in distance education courses (using linear regression) [190]; for predicting when a student will get a question correct and association rules to guide a search process to find transfer models to predict a student's success (using logistic regression) [89]; to predict the probability a student has of giving the correct answer to a problem in an ITS (using a robust Ridge regression algorithm) [61]; for predicting end-of-year accountability assessment scores (using linear regression) [7], to predict a student's test score (using stepwise regression) [80] and to predict the probability that the student's next response has of being correct (using linear regression) [31].

Finally, correlation analyses have been applied together to predict web-student performance in on-line classes [277]; to predict a student's final exam score in online tutoring [209] and for predicting high school students' probabilities of success in university [175].

E. *Student Modeling*

The objective of student modeling is to develop cognitive models of human users/students, including a modeling of their skills and declarative knowledge. Data mining has been applied to automatically consider user characteristics (motivation, satisfaction, learning styles, affective status, etc.) and learning behavior in order to automate the construction of student models [90]. Different DM techniques and algorithms have been used for this task (mainly, Bayesian networks).

Several data mining algorithms (Naïve Bayes, Bayes net, support vector machines, logistic regression and decision trees) have been compared to detect student mental models in intelligent tutoring systems [236]. Unsupervised (clustering) and supervised (classification) machine learning have been proposed to reduce development costs in building user models and to facilitate transferability in intelligent learning environments [4]. Clustering and classification of learning variables have been used to measure the online learner's motivation [117].

Bayesian networks have been used to make predictions about student knowledge, i.e. the probability that student has of knowing a skill at a given time through cognitive tutors [18]; to detect students' learning styles in a web-based education system [92]; to predict whether a student will answer a problem correctly [136]; to model a student's changing state of knowledge during skill acquisition in ITS [47]; to infer unobservable learning variables from students' help-seeking behavior in a web-based tutoring system [10] and for knowledge tracing in order to verify the impact of self-discipline on students' knowledge and learning [99].

Sequential pattern mining has been used to automatically acquire the knowledge to construct student models [9]; to identify meaningful user characteristics and to update the user model to reflect newly gained knowledge [6] and for predicting students' intermediate mental steps in sequences of actions stored by - learning environments based on problem solving [220].

Association rule algorithms have been applied for personality mining based on web-based education models in order to deduce learners' personality characteristics [122] and for student modeling in intelligent tutoring systems [170].

Other DM techniques and models have also been used for student modeling. A logistic regression model has been used to construct transfer models (to accurately predict the level at which a student represents knowledge) [84]. A learning agent that models student behaviors using linear regression has been constructed in order to predict the probability that the student's next response has of being correct [31]. Inductive logic programming and a profile extractor system (using numeric algorithms) have been developed to induce student profiles in e-learning systems [157]. The Markov decision process has been proposed to automatically create student models by generating hints for an intelligent tutoring that learns [26]. Fuzzy techniques have used student models in web-based learning environments in order to generate advice for the teachers [146]. A dynamic learning response model has been developed for inferring, testing and verifying student learning models on an adaptive learning website [127]. Bootstrapping novice data can create an initial skeletal model of a tutor from log data collected from actual use of the tool by students [176]. A collaborative-based data mining approach has been developed for diagnostic and predictive student modeling purposes in integrated learning environments [153]. Multiple correspondence analysis and cross-validation by correlation analysis have been applied to identify learning styles in ILS (Index of Learning Styles) questionnaires [272]. The Q-matrix method has been used to create concept models that represent relationships between concepts and questions, and to group student test question responses according to concepts [25]. An algorithm to estimate Dirichlet priors has been developed to produce model parameters that provide a more plausible picture of student knowledge [215]. Self-organizing maps and principal component analysis have been applied for predictive and compositional modeling of the student profile [152]. A clustering algorithm (K-means) has been developed to model student behavior with a very small set of parameters without compromising the behavior of the system [219].

F. Detecting undesirable student behaviors

The objective of detecting undesirable student behavior is to discover/detect those students who have some type of problem or unusual behavior such as: erroneous actions, low motivation, playing games, misuse, cheating, dropping out, academic failure, etc. Several DM techniques (mainly, classification and clustering) have been used to reveal these types of students in order to provide them with appropriate help in plenty of time.

Several of the classification algorithms that have been used to detect problematic student behavior are decision tree neural networks, naïve Bayes, instance-based learning, logistic regression and support vector machines for predicting/preventing student dropout [147]; feed-forward neural networks, support vector machines and a probabilistic ensemble simplified fuzzy ARTMAP algorithm to predict dropouts in e-learning courses [158]; Bayesian nets, logistic regression, simple logic classification, instance based classification, attribute selected classification, bagging, classification via regression and decision trees for engagement prediction [64]; decision tree, Bayesian classifiers, logistic models, the rule-based learner and random forest to detect/predict first year student drop out [66]; paired t-test for grouping students by common misconceptions (hint-driven learners and failure-driven learners) [289]; C4.5 decision tree algorithm for detecting any potential symptoms of low performance in e-learning courses [41]; decision trees to identify students with little motivation [63]; decision trees for detection of irregularities and deviations in the learners' actions in an interactive learning environment [189]; and the J48 decision tree algorithm and FarthestFirst clustering algorithm for predicting, understanding and preventing academic failure (exam failure) among university students [42].

Different types of clustering also used to carry this task out are: Kohonen nets to detect students that cheat in online assessments [43]; outlier detection to uncover atypical student behavior [267]; an outlier detection method using Bayesian predictive distribution to detect learners' irregular learning [265]; a constrained mixture of student t-distribution and generative topographic mapping to detect atypical student behavior (outliers) [59] and an augmented version of the Levenshtein distance algorithm to identify novice errors and error paths [267].

Finally, other DM techniques and models used for this task are, for example: association rule mining for selecting weak students for remedial classes [165], to send warning messages to students with unusual learning behavior in an adaptive educational hypermedia system [135], and to construct concept-effect relationships for diagnosing student learning problems [128]; a latent response model to identify if students are playing with the system (to detect student misuse) in a way that would lead to poor learning [15] and to automatically detect when a student is off-task in a cognitive tutor [16]; Bayesian networks to predict the need for help in an interactive learning environment [171]; stepwise regression to detect misplay and look for sources of error in the prediction of student test scores [80]; human reliability analysis to infer the underlying causes that lead to the production of trainee errors in a virtual environment [74] and Markov chain analysis to identify and classify common student errors and technical problems in order to prevent them from occurring in the future [111].

G. Grouping students

The objective is to create groups of students according to their customized features, personal characteristics, etc. Then, the clusters/groups of students obtained can be used by the instructor/developer to build a personalized learning system, to promote effective group learning, to provide adaptive contents, etc. The DM techniques used in this task are classification (supervised learning) and clustering (unsupervised learning). Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster have some points in common [231].

Different clustering algorithms have been used to group students, such as: hierarchical agglomerative clustering, K-means and model based clustering to identify groups of students with similar skill profiles [14]; a clustering algorithm based on large generalized sequences to find groups of students with similar learning characteristics based on their traversal path patterns and the content of each page they have visited [258]; model-based clustering to automatically discover useful groups from LMS data to obtain profiles of student behavior [256]; a hierarchical clustering algorithm for user modeling (learning styles) in intelligent e-learning systems in order to group students according to their individual learning style preferences [296]; discriminating features and external profiling features (pass/fail) to support teachers in collaborative student modeling [91]; an improvement in the matrix-based clustering method for grouping learners by characteristics in e-e-learning [297]; a fuzzy clustering algorithm to find interested groups of learners according to their personality and learning strategy data collected from an online course [261]; a hybrid method of clustering and Bayesian networks to group students according to their skills [107]; a K-means clustering algorithm for effectively grouping students who demonstrate similar learning portfolios (students' assignment scores, exam scores and online learning records) [51]; an Expectation-Maximization algorithm to form heterogeneous groups according to student skills [190]; a K-means clustering algorithm to discover interesting patterns that characterize the work of stronger and weaker students [211]; a conditional subspace clustering algorithm to identify skills which differentiate students [196]; a two-step cluster analysis to classify how students organize personal information spaces (piling, one-folder, small-folders and big-folder filing) [110]; hierarchical cluster analysis to establish the proportion of students who get an exercise wrong or right [24]; a genetic clustering algorithm to solve the problem of allocating new students (which places new students into classes so that the gaps between learning levels in each class is minimum and the number of students in each class does not exceed the limit) [306].

Several classification algorithms have been applied in order to group students, such as: discriminant analysis, neural networks, random forests and decision trees for classifying university students into three groups (low-risk, medium-risk and high-risk of failing) [254]; classification and regression tree, chi-squared automatic interaction detection and C4.5 algorithm for the automatic identification of the students' cognitive styles [155]; a classification and regression tree to create a decision tree model to illustrate a user´s learning behavior in order to analyze it according to different cognitive style groups [153]; a hidden Markov-model-based classification approach to characterize different types of users through their navigation or content access patterns [86]; decision trees for classifying students according to their accumulated knowledge in e-learning systems [181]; C4.5 decision tree algorithm for discovering potential student groups with similar characteristics who will react to a particular strategy [49]; Naïve Bayes classifier to classify learning styles that describe learning behavior and educational content [140]; genetic algorithms for grouping students according to their profiles in a peer review content [65]; classification trees and multivariate adaptive regression to identify those students who tend to take online courses and those who do not [292]; decision tree and support vector machine for assessing an activity by more than one lecturer using a pair-wise learning model [212]; a classification algorithm for speech act patterns to assess participants' roles and identify discussion threads [143] and K-nearest neighbor (K-NN) classification combined with genetic algorithms to identify and classify student learning styles [48].

H. Social network analysis

Social Networks Analysis (SNA), or structural analysis, aims at studying relationships between individuals, instead of individual attributes or properties. A social network is considered to be a group of people, an organization or social individuals who are connected by social relationships like friendship, cooperative relations, or informative exchange [88]. Different DM techniques have been used to mine social networks in educational environments, but collaborative filtering is the most common. Collaborative filtering or social filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting taste preferences from many users (collaborating) [115]. Collaborative filtering systems can produce personal recommendations by computing the similarity between students' preferences, so this task is directly related to the previous task of recommendations for students (see Section F).

Collaborative filtering has been used for context-aware learning object recommendation lists [156]; to make a recommendation for a learner about what he/she should learn before taking the next step [302]; for developing a personal recommender system for learners in lifelong learning networks [71]; to build a resource recommendation system based on connecting to similar e-learning [287]; for recommending relevant links to the active learner [149]; to develop an e-learning recommendation service system [159] and to find relevant content on the web, personalizing and adapting this content to learners [259].

There are some other DM techniques that have been applied to analyze social networks. Mining interactive social networks has been proposed for recommending appropriate learning partners in a web-based cooperative learning environment [53]. Social navigation support and various machine learning methods have been used in a course recommendation system in order to make relevant course choices based on students' assessment of course relevance for their career goals [77]. Social network analysis techniques and mining data produced by students involved in communication through forum-like tools have been suggested to help reveal aspects of their communication [235]. Data mining and social networks have been used to analyze the structure and content of educative on-line communities [218]. Social network analysis has been proposed to detect patterns of academic collaboration in order to aid decision makers in organizations to take specific actions depending on the patterns [192]. Analysis of social communicative categories has been suggested to distinguish between a variety of speech acts (informing belief, disagreeing with concepts, offering collaborative acts, and insulting) [208]. Visualizing and clustering on discussion forum graphs have been applied as social network analysis to measure the cohesion of small groups in collaborative distance-learning [233].

I. *Developing concept maps*

The objective of constructing concept maps is to help instructors/educators in the automatic process of developing/constructing concept maps. A concept map is a conceptual graph that shows relationships between concepts and expresses the hierarchal structure of knowledge [195]. Some DM techniques (mainly, association rules and text mining) have been used to construct concept maps.

Association rule mining has been used to automatically construct concept maps guided by learners' historical testing records [264]; to discover concept-effect relationships for diagnosing the learning problems of students [128] and for conceptual diagnosis of e-learning through automatically constructed concept maps that enable teachers to overcome the learning barrier and misconceptions of learners [154].

Text mining has been applied to automatically construct concept maps from academic articles in the e-learning domain [52]; to formulate concept maps from online discussion boards using fuzzy ontology [151]; to find relationships between text documents and construct document index graphs [109] and to explore cognitive concept-map differences in instructional outcomes [121].

Finally, a specific concept-map algorithm has been created to automatically organize knowledge points and map them [245]; a method of automatic concept relationship discovery for an adaptive e-course has been developed to help teachers to author overall automation [247] and a multi-expert e-training course design model has been developed by concept map generation in order to help the experts to organize their domain knowledge [58].

J. *Constructing courseware*

The objective of constructing courseware is to help instructors and developers to carry out the construction/development process of courseware and learning contents automatically. On the other hand, it also tries to promote the reuse/exchange of existing learning resources among different users and systems.

Different DM techniques and models have been used to develop courseware. The clustering of students and naïve algorithms have been proposed to construct personalized courseware by building a personalized web tutor tree [257]. Rough set theory and clustering concept hierarchy have been used to construct e-learning FAQ retrieval infrastructures [56]. Multilingual knowledge-discovery technique processing has been combined with Adaptive Hypermedia techniques to automatically create on-line information systems from linear texts in electronic format, such as textbooks [3]. Argument mining has been proposed to support argument construction for agents and intelligent tutoring systems using different mining techniques [1].

Several DM techniques have been applied to reuse learning resources. Hybrid unsupervised data mining techniques have been employed to facilitate Learning Object (LO) reuse and retrieval from the Web or from different LO repositories [144]. Valuable information can be found by mining metadata from educational resources (ontology of pedagogical objects) which helps data mining to retrieve more precise information for content re-use and exchange [177]. The automatic classification of web documents in a hierarchy of concepts based on Naïve Bayes has been suggested for the indexing and reuse of learning resources [237]. Profile analysis based on collaborative filtering has been used to search learning objects and rank search results according to the predicted level of user interest [199]. Mining educational multimedia presentations has been used to establish explicit relationships among the data related to interactivity (links and actions) and to help predict interactive properties in the multimedia presentations [13].

K. *Planning and scheduling*

The objective of planning and scheduling is to enhance the traditional educational process by planning future courses, helping with student course scheduling, planning resource allocation, helping in the admission and counseling processes, developing curriculum, etc. Different DM techniques have been used for this task (mainly, association rules).

Classification, categorization, estimation and visualization have been compared in higher education for different objectives, such as academic planning, predicting alumni pledges and creating meaningful learning outcome typologies [164]. Decision trees, link analysis and decision forests have been used in course planning to analyze enrollees' course preferences and course completion rates in extension education courses [120]. Classification, prediction, association rule analysis, clustering, etc. have been compared to discover new explicit knowledge which could be useful in the decision making process in higher learning institutions [68]. Educational training courses have been planned through the use of cluster analysis, decision trees and back-propagation neural networks in order to find the correlation between the course classifications of educational training [123]. Decision trees and Bayesian models have been proposed to help management institutes explore the probable effects of changes in recruitments, admissions and courses [217].

Association rule mining has been used to provide new, important and therefore demand-oriented impulses for the development of new bachelor and master courses [239]. Curriculum revision has been done by association rule mining in order to identify and understand whether curriculum revisions can affect students in a university [36]. A decisional tool (based on association rule mining) has been constructed to help make decisions on how to improve the quality of the service provided by the university based on students' success and failure rates [241]. Association rule mining and genetic algorithms have been applied to an automatic course scheduling system to produce the course timetables that best suit student and teacher needs [280].

Finally, a regression model has been developed to predict the likelihood a specific undergraduate applicant has of matriculating if admitted [141]; several clustering algorithms (self-organizing map networks, K-means and kth-nearest neighbor) have been used as a decision support in selecting AACSB (Association of Advance Collegiate Schools of Business) peer schools [142].

III. FUTURE WORK AND RESEARCH LINES

Although there is a lot of future work to be considered in EDM, we indicate in continuation what are arguably the most interesting and influential among them. In fact, a few initial studies on some of these points have already begun to appear.

- EDM tools have to be designed to be easier for educators or non-expert users in data mining. Data mining tools are normally 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 may want to do. For example, on the one hand, users have to select the specific DM method/algorithm they want to apply/use from the wide range of methods/algorithms available on DM. On the other hand, most of the data mining algorithms need to be configured before they are executed. Users have to provide appropriate values for the parameters in advance in order to obtain good results/models and so, the user must possess a certain amount of expertise in order to find the right settings. One possible solution is the development of wizard tools that use a default algorithm for each task and parameter-free data mining algorithms to simplify the configuration and execution for non-expert users. EDM tools must also have a more intuitive interface that is easy to use and with good visualization facilities to make their results meaningful to educators and e-learning designers [94]. It is also very important to develop specific preprocessing tools in order to automate and facilitate all the preprocessing functions or tasks that EDM users currently must do manually.
- Integration with the e-learning system. The data mining tool has to be integrated into the e-learning environment as one more traditional authoring tool (course creator, test creator, report tools, etc.). All data mining tasks (preprocessing, data mining and postprocessing) must be carried out in a single application with a similar interface. In this way, EDM tools will be more widely used by educators, and feedback and results obtained with data mining techniques could be easily and directly applied to the e-learning environment using an iterative evaluation process [226].
- Standardization of data and models. Current tools for mining data pertaining to a specific course/framework may be useful only to their developers. There are no general tools or re-using tools that can be applied to any educational system. So, a standardization of input data and output model are needed, as along with preprocessing, discovering and postprocessing tasks. Some authors [245] have proposed using XML as data specification. Other authors [269] have used PMML (Predictive Modeling Markup Language) that is the leading standard for statistical and data mining models. But it is also necessary to incorporate domain knowledge and semantics using ontology specification languages, such as OWL (Ontology Web Language) and RDF (Resource Description Framework); and standard metadata for e-learning such as SCORM (Sharable Content Object Reference Model). In this line, currently, there is only one public educational data repository, the PSLC DataShop [145], which provides a lot of educational data sets and also facilitates analysis. However, all this log data is obtained from Intelligent Tutoring Systems, so it is necessary to have more public datasets from other types of educational environments as well. In this way, specific educational benchmark datasets could be used to compare/evaluate different data mining algorithms.
- Traditional mining algorithms need to be tuned to take into account the educational context. Data mining techniques must use semantic information when applied to educational data. This shows the need for more effective mining tools that integrate educational domain knowledge into data mining algorithms. For example, some authors [131] have proposed specific usage tracking language (UTL) to describe the track semantics recorded by a Learning Management system and to link them to the need for observation defined in a predictive scenario. Education-specific mining techniques can greatly

improve instructional design and pedagogical decisions, and the aim of the semantic web is to facilitate data management in educational environments.

IV. Conclusion

This paper is a review of the state of the art with respect to EDM and surveys the most relevant work in this area to date. Each study has been classified, not only by the type of data and DM techniques used, but also and more importantly, by the type of educational task that they resolve. EDM has been introduced as an up and coming research area related to several well-established areas of research including e-learning, adaptive hypermedia, intelligent tutoring systems, web mining, data mining, etc. We have seen how fast EDM is growing as reflected in the increasing number of contributions published every year in International Conferences and Journals, and the number of specific tools specially developed for applying data mining algorithms in educational data/environments. So, it could be said that EDM is now approaching its adolescence, that is, it is no longer in its early days but is not yet a mature area. In fact, we have described some interesting future lines but for it to become a more mature area it is also necessary for researchers to develop more unified and collaborative studies instead of the current plethora of multiple individual proposals and lines. Thus, the full integration of data mining in the educational environment will become a reality, and fully operative implementations (both commercial and free) could be made available not only for researchers and developers but also for external users.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the Spanish department of Research under TIN2008-06681-C06-03 and P08-TIC-3720 Projects and FEDER funds.

References

[1] Abbas, S., Sawamura, H. (2008). A First Step towards Argument Mining and Its Use in Arguing Agents and ITS. In International Conference on Knowledge-Based intelligent information and Engineering Systems, Zagreb, Croatia, 149-157.

[2] Abel, F., Bittencourt, I. I., Henze, N., Krause, D., Vassileva, J. (2008). A Rule-Based Recommender System for Online Discussion Forums. In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Hannover, Germany,12-21.

[3] Alfonseca, E., Rodriguez, P., Perez, D. (2007). An approach for automatic generation of adaptive hypermedia in education with multilingual knowledge discovery techniques. Computers & Education Journal, 49, 2, 495-513.

[4] Amershi, S., Conati, C., (2009). Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments. Journal of Educational Data Mining, 1, 1, 18-71.

[5] Anaya, A., Boticario, J. (2009). A data mining approach to reveal representative collaboration indicators in open collaboration frameworks. In International Conference on Educational Data Mining, Cordoba, Spain, 210-218.

[6] Andrejko. A., Barla, M., Bielikova, M., Tvarozek, M. (2007). User Characteristics Acquisition from Logs with Semantics. In International Conference on Information System Implementation and Modeling, Czech Republic, 103-110.

[7] Anozie, N., Junker, B.W. (2006). Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring system. In Educational Data Mining AAAI Workshop, California, 1-6.

[8] Arnold, A., Scheines, R., Beck, J.E., Jerome, B. (2005). Time and Attention: Students, Sessions, and Tasks. In AAAI2005 Workshop on Educational Data Mining, Pittsburgh, 62-66.

[9] Antunes, C. 2008. Acquiring background knowledge for intelligent tutoring systems. In International Conference on Educational Data Mining, Montreal, Canada, 18-27.

[10] Arroyo, I., Murray, T., Woolf, B.P. (2004). Inferring Unobservable Learning Variables from Students' Help Seeking Behavior. In International Conference on Intelligent Tutoring System, Brazil, 782-784.

[11] Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., Voyiatzaki, E. (2005). Why logging of fingertip actions is not enough for analysis of learning activities. In Workshop on Usage analysis in learning systems, AIED Conference, Amsterdam, 1-8.

[12] Ayers E., Junker B.W. (2006). Do skills combine additively to predict task difficulty in eighth grade mathematics? In AAAI Workshop on Educational Data Mining: Menlo Park, 14-20.

[13] Bari, M. Lavoie, B. (2007). Predicting interactive properties by mining educational multimedia presentations. In International Conference on Information and Communications Technology, 231-234.

[14] Ayers, E., Nugent, R., Dean, N. (2009). A Comparison of Student Skill Knowledge Estimates. In International Conference On Educational Data Mining, Cordoba, Spain, 1-10.

[15] Baker, R., Corbett, A., Koedinger, K. (2004). Detecting student misuse of intelligent tutoring systems. In International Conference on Intelligent Tutoring Systems, Alagoas, Brazil, 531–540.

[16] Baker, R., (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In Conference on Human Factors in Computing Systems, San Jose, California, 1059-1068.

[17] Baker, R., Beck, J.E., Berendt, B., Kroner, A., Menasalvas, E., Weibelzahl, S. (2007). Track on Educational Data Mining, at the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling. Corfu, Greece.

[18] Baker, R., Corbett, A.T., Aleven, V. (2008). Improving contextual models of guessing and slipping with a truncated training set. In International Conference on Educational Data Mining, Montreal, Canada, 67-76.

[19] Baker, R., Barnes, T., Beck, J.E. (2008). Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings. Montreal, Quebec, Canada.

[20] Baker, R., Yacef, K. (2009) The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1, 1, 3-17.

[21] Baker, R (2010). Data Mining for Education. To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

[22] Baker, R., Merceron, A., Pavilk, P.I. (2010). Educational Data Mining 1010: 3st International Conference on Educational Data Mining, Proceedings, Pittsburgh, USA.

[23] Ba-omar, Petrounias, I., Anwar, F. (2007). A framework for using web usage mining for personalise e-learning. In International Conference on Advanced Learning Technologies, Niigata, Japan, 937–938.

[24] Barker-Plummer, D., Cox, R., Dale, R. (2009). Dimensions of difficulty in translating natural language into fist order logic. In International Conference on Educational Data Mining, Cordoba, Spain, 220-228.

[25] Barnes, T. (2005). The q-matrix method: Mining student response data for knowledge. In Proceedings of the AAAI-2005 Workshop on Educational Data Mining, Pittsburgh, PA, 1-8.

[26] Barnes, T., Stamper, J. (2008). Toward Automatic Hint Generation for Logic Proof Tutoring Using Historical Student Data. In International Conference on intelligent Tutoring Systems, Montreal, Canada, 373-382.

[27] Barnes, T., Desmarais, M., Romero, C., Ventura, S. (2009). Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings. Cordoba, Spain.
[28] Baruque, C. B., Amaral, M.A., Barcellos, A., Da Silva Freitas, J.C., Longo, C. J. (2007). Analysing users' access logs in Moodle to improve e learning. In Euro American Conference on Telematics and information Systems, Faro, Portugal, 1-4.
[29] Beal, C. R. and Cohen, P. R. (2008). Temporal Data Mining for Educational Applications. In Proceedings of the 10th Pacific Rim international Conference on Artificial intelligence: Trends in Artificial intelligence, Hanoi, Vietnam, 66-77.
[30] Beck, J.E. 2000. Workshop on Applying Machine Learning to ITS Design/Construction at the 5th International Conference on Intelligent Tutoring Systems, ITS2000, Montreal, Canada.
[31] Beck, J.E., Woolf, B.P. (2000). High-level student modeling with machine learning. In Fifth International Conference on Intelligent Tutoring Systems, Alagoas, Brazil, 584-593.
[32] Beck, J.E., Baker, R., Corbett, A.T., Kay, J., Litman, D.J., Mitrovic, T., Ritter, S. (2004). Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes at 7th International Conference, Alagoas, Brazil.
[33] Beck, J.E. (2005). Workshop on Educational Data Mining at the 20th National Conference on Artificial Intelligence, AAAI2005. Pittsburgh, USA.
[34] Beck, J.E., Aimeur, E., Barnes, T. (2006). Workshop on Educational Data Mining at the 21st National Conference on Artificial Intelligence, AAAI2006. Boston, USA.
[35] Beck, J.E., Pechenizkiy, M., Calders, T., Viola, S.R. (2007). Workshop on Educational Data Mining at the 7th IEEE International Conference on Advanced Learning Technologies. Niigata, Japan.
[36] Becker, K., Ghedini, C., Terra, E. (2000). Using kdd to analyze the impact of curriculum revisions in a Brazilian university. In Eleventh international conference on data engineering. Orlando, 412–419.
[37] Bellaachia, A., Vommina, E. (2006). MINEL: A framework for mining e-learning logs. In Fifth IASTED International Conference on Web-based Education, Mexico, 259-263.
[38] Ben-naim, D., Marcus, N., Bain, M. (2008). Visualization and Analysis of Student Interaction in an Adaptive Exploratory Learning Environment. In Int. Workshop in Intelligent Support for Exploratory Environments in the European Conference on Technology Enhanced Learning, Maastricht, 1-10.
[39] Ben-naim, D., Bain, M., Marcus, N. (2009). A User-Driven and Data-Driven Approach for Supporting Teachers in Reflection and Adaptation of Adaptive Tutorials. In International Conference on Educational Data Mining, Cordoba, Spain, 21-30.
[40] Burr, L., Spennemann, D.H. (2004). Pattern of user behavior in university online forums. In International Journal of Instructional Technology and Distance Learning, 1,10 11-28.
[41] Bravo, J., Ortigosa, A. (2009). Detecting Symptoms of Low Performance Using Production Rules. In International Conference on Educational Data Mining, Cordoba, Spain,
[42] Bresfelean, V.P., Bresfelean, M., Ghisoiu, N. (2008). Determining students' academic failure profile founded on data mining methods. In International Conference on Information Technology Interfaces, Croatia, 317–322.
[43] Burlak, G., Muñoz, J., Ochoa, A., Hernández, J.A. (2006). Detecting Cheats In Online Student Assessments Using Data Mining. In International Conference on Data Mining, Las Vegas, 204-210.
[44] Cakir, M., Xhafa, F., Zhou, N., Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. In International conference on AI in Education, Amsterdam, 121-127.
[45] Chan, C.C. (2007). A Framework for Assessing Usage of Web-Based e-Learning Systems. In International Conference on innovative Computing, Information and Control, Washington, DC, 147- 151.
[46] Chanchary, F. H., Haque, I., Khalid, M. S. (2008). Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment. In International Workshop on Knowledge Discovery and Data Mining, Washington, 249-253.
[47] Chang, K.M., Beck, J.E., Mostow, J., Corbett, A. (2006). A Bayes Net Toolkit for Student Modeling in Intelligent Tutoring Systems. In International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan, 104-113.
[48] Chang, Y.C, Kao, W.Y., Chu, C.P., Chiu, C.H. (2009). A learning style classification mechanism for e-learning. In Computer & Education Journal, 53,2, 273-285.
[49] Chen, G., Liu, C., Ou, K., Liu, B. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. Journal of Educational Computing Research 23,3, 305–332.
[50] Chen, C., Duh, L., Liu, C. (2004). A Personalized Courseware Recommendation System Based on Fuzzy Item Response Theory. In IEEE international Conference on E-Technology, E-Commerce and E-Service, Washington, DC, 305-308.
[51] Chen, c., Chen, M., Li, Y. (2007). Mining key formative assessment rules based on learner portfiles for web-based learning systems. In IEEE International Conference on Advanced Learning Technologies, Japan, 1-5.
[52] Chen, N.S., Kinshuk, Wei, C.W., Chen, H.J. (2008). Mining e-learning domain concept map from academic articles. In Computers & Education Journal, 50, 1009-1021.
[53] Chen, C.H., Hong, C.M., Chang, C.C. (2008). Mining interactive social network for recommending appropriate learning partners in a Web-based cooperative learning environment. In IEEE Conference on Cybernetics and Intelligent Systems, Chengdu, 642-647.
[54] Chen, Y., Weng, C. (2009). Mining fuzzy association rules from questionnaire data. In Knowledge-Based Systems Journal, 22,1, 46-56.
[55] Chen, C., Chen, M. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. Computer & Educaction Journal. 52,1, 256-273.
[56] Chiu, D.Y., Pan, Y.C., Chang, W.C. (2008). Using rough set theory to construct e-learning faq retrieval infrastructure. In IEEE Ubi-Media Computing Conference. Lanzhou, 547-552.
[57] Chu, H.C., Hwang, G.J., Tseng, J.C.R., Hwang, G.H. (2006). A computerized approach to diagnosing student learning problems in health education. In Asian Journal of Health and Information Sciences, 1,1, 43-60.
[58] Chu, H.C., Hwang, G.J., Wu, P.H., Chen, J.M. (2007). A Computer-assisted Collaborative Approach for E-training Course Design. In IEEE Conference on Advanced Learning Technologies, Niigata, Japan, 36-40.
[59] Castro, F., Vellido, A., Nebot, A., Minguillon, J. (2005). Detecting atypical student behaviour on an e-learning system. In Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, Granda, Spain, 153-160.
[60] Castro, F., Vellido, A., Nebot, A. Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems. In: Jain, L.C., Tedman, R. and Tedman, D. (eds.) Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62, Springer-Verlag, 183-221.
[61] Cetintas, A., Si, L., Xin, Y.P., Hord, C. (2009). Predicting correctness of problem solving from low-level log data in intelligent tutoring systems. In International Conference on Educational Data Mining, Cordoba, Spain, 230-238.
[62] Choquet, C., Luengo, V., Yacef, K. (2005). Workshop on Usage Analysis in Learning Systems at Artificial Intelligence in Education Conference AIED-2005. Amsterdam, The Netherlands.
[63] Cocea, M., Weibelzahl, S. (2006). Can Log Files Analysis Estimate Learners' Level of Motivation? In Workshop week Lernen - Wissensentdeckung - Adaptivität, Hildesheim, 32-35.
[64] Cocea., M., Weibelzahl., S. (2007). Cross-system validation of engagement prediction from log files. In International Conference on Conference on Technology Enhanced Learning, Crete, Greece, 14-25.
[65] Crespo, R. M., Pardo, A., Pérez, J.P., Kloos, C.D. (2005). An algorithm for peer review matching using student profiles based on fuzzy classification and genetic algorithms. In International Conference on innovations in Applied Artificial Intelligence, Bari, Italy, 685-694.

[66] Dekker, G.W., Pechenizkiy, M., Vleeshouwers, J.M. (2009). Predicting Students Drop Out: A Case Study. In International Conference on Educational Data Mining, Cordoba, Spain, 41-50.
[67] Delgado, M., Gibaja, E., Pegalajar, M.C., Pérez, O. (2006). Predicting Students' Marks from. Moodle Logs using Neural Network Models. In International Conference on Current Developments in Technology-Assisted Education, Sevilla, Spain, 586-590.
[68] Delavari, N., Phon-amnuaisuk, S., Beikzadeh, M. (2008). Data Mining Application in Higher Learning Institutions. In Informatics in Education Journal, 7,1, 31-54.
[69] Desmarais, M.C., Gagnon, M., Meshkinfram, P. (2006). Bayesian Student Models Based on Item to Item Knowledge Structures In Conference on Technology Enhanced Learning, Crete, Greece, 1-10.
[70] Dong, G., Pei, J. (2007). Sequence Data Mining. Springer.
[71] Drachsler, H., Hummel, H. G., Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. In International Journal of Learning Technology. 3,4, 404-423.
[72] Draper, N.R., Smith, H. (1998). Applied Regression Analysis. Wiley.
[73] Dringus, L.P., Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. In Computer & Education Journal. 45, 1, 141-160.
[74] El-kechaï, N., Després, C. (2007). Proposing the underlying causes that lead to the trainnee's erroneous actions to the trainer. In European Conference on Technology-Enhanced Learning, Crete, Creece, 41-55.
[75] Espejo, P., Ventura, S., Herrera, F. (2010) A Survey on the Application of Genetic Programming to Classification. IEEE Transactions on Systems, Man, and Cybernetics-Part C. 40, 2, 121-144.
[76] Etchells, T.A., Nebot, A., Vellido, A., Lisboa, P.J.G., Mugica, F. (2006). Learning what is important: feature selection and rule extraction in a virtual course. In European Symposium on Artificial Neural Networks, Bruseles, Belgium, 401-406.
[77] Farzan R., Brusilovsky P. (2006). Social Navigation Support in a Course Recommendation System. In proceedings of 4th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems. Dublin, 91-100.
[78] Fayyad, U. Piatetsky-shapiro G., Smyth. P. (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, 17, 37-54.
[79] Fausett, L.V., Elwasif, W. (1994). Predicting performance from test scores using backpropagation and counterpropagation. In IEEE World Congress on Computational Intelligence, Paris, France, 3398–3402.
[80] Feng, M., Heffernan, N., Koedinger, K. (2005). Looking for sources of error in predicting student's knowledge. In: AAAI'05 workshop on Educational Data Mining, 1-8.
[81] Feng, M., Heffernan, N.T. (2005). Informing Teachers Live about Student Learning: Reporting in the Assistment System. In Conference on Artificial Intelligence in Education 2005 Workshop on Usage Analysis in Learning Systems, Amsterdam, 1-8.
[82] Feng, M., Heffernan, N. (2006). Informing Teachers Live about Student Learning: Reporting in the Assistment System. Technology, Instruction, Cognition, and Learning Journal, 3,1-8.
[83] Feng, M., Heffernan, N.T, Mani, M., Heffernan C. (2006). Using Mixed-Effects Modeling to Compare Different Grain-Sized Skill Models. In Workshop on Educational Data Mining, Menlo Park, CA,. 57-66.
[84] Feng, M., Beck, J. (2009). Back to the future: a non-automated method of constructing transfer models. In International Conference on Educational Data Mining, Cordoba, Spain, 240-248.
[85] Feng., M. Beck., J.E., Heffernan, N.T. (2009). Using Learning Decomposition and Bootstrapping with Randomization to Compare the Impact of Different Educational Interventions on Learning. In International Conference On Educational Data Mining, Cordoba, Spain, 51-60.
[86] Fok, A.W.P., Wong, H.S., Chen, Y.S. (2005). Hidden Markov model based characterization of content access patterns in an e-learning environment. In IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands, 201 – 204.
[87] Freedman, D., Purves, R., Pisani, R. (2007). Statistics, 4 th Edition. W.W. Norton & Co.
[88] Freeman, L. (2006). The Development of Social Network Analysis. Empirical Press.
[89] Freyberger, J., Heffernan, N.T., Ruiz, C. (2004). Using Association Rules to Guide a Search for Best Fitting Transfer Models of Student Learning. In Workshop Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, Alagoas, Brazil, 1-10.
[90] Frias-Martinez, E. Chen, S., Liu, X. Survey of Data Mining Approaches to User Modeling for Adaptive Hypermedia. IEEE Transactions on Systems, Man, and Cybernetics-Part C. 36, 6, 734-749.
[91] Gaudioso, E., Montero, M., Talavera, L., Hernandez-del-olmo, F. (2009). Supporting teachers in collaborative student modeling: a framework and an implementation. In Expert System with Applications, 36, 2260-2265.
[92] Garcia, P., Amandi, A., Schiaffino, S., Campo, M. (2007). Evaluating bayesian networks' precision for detecting student's learning styles. In Computer & Education Journal. 49, 794-808.
[93] Garcia, E., Romero, C., Ventura, S., Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 19, 99–132.
[94] Garcia, E., Romero, C., Ventura, S., Castro, C. (2009b). Collaborative Data Mining Tool for Education. In International Conference on Educational Data Mining, Cordoba, Spain, 299-306.
[95] Gedeon, T.D., Turner, H.S. (1993). Explaining student grades predicted by a neural network. In International conference on Neural Networks, Nagoya, 609-612.
[96] Gibbs, J., Rice, M. (2003). Evaluating student behavior on instructional Web sites using web server logs. In Ninth Sloan-C International Conference on Online Learning, Orlando, 1-3.
[97] Girones, M., Fernandez, T.A. (2006) Ariadne, a guiding thread in the learning process's labyrinth. In International Conference on Current Developments in Technology-Assisted Education, Sevilla, 287-290.
[98] Golding, P., Donalson, O. (2006. Predicting Academic Performance. In Frontiers in Education Conference. San Diego, California, 21-26.
[99] Gong, Y., Rai, D., Beck, J.E., Heffernan, N.T. (2009). Does Self-Discipline impact students' knowledge and learning?. In International Conference On Educational Data Mining, Cordoba, Spain, 61-70.
[100]Grob, H.L., Bensberg, F., Kaderali, F. (2004). Controlling Open Source Intermediaries – a Web Log Mining Approach. In International Conference on Information Technology Interfaces, Zagreb, 233-242.
[101]Guo, Q., Zhang, M. (2009). Implement web learning environment based on data mining. In Knowledge-Based Systems Journal, 22, 439-442.
[102]Ha, S., Bae, S., Park, S. (2000). Web Mining for Distance Education. In IEEE International Conference on Management of Innovation and Technology, Singapore, 715–719.
[103]Haddawy, P., Thi, N., Hien, T.N. (2007). A decision support system for evaluating international student applications. In Frontiers In Education Conference, Milwaukee, 1-4.
[104]Hadwin, A. F. Nesbit, J. C. Jamieson-Noel, D. Code, J. Winne, P. H. (2007). Examining trace data to explore self-regulated learning. In Metacognition and Learning Journal, Volumen 2, Number 2-3, 107-124.
[105]Han, J., Kamber, M. (2006). Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers.
[106]Hanna, M. (2004). Data mining in the e-learning domain. In Campus-Wide Information Systems, Volume 21, Number 1, 29-34.
[107]Hämäläinen, W., Suhonen, J., Sutinen, E., Toivonen, H. (2004). Data mining in personalizing distance education courses. In World Conference on Open Learning and Distance Education, Hong Kong, 1-11.
[108]Hämäläinen, W., Vinni, M. (2006). Comparison of machine learning methods for intelligent tutoring systems. In international conference in intelligent tutoring systems, Taiwan, 525-534.
[109]Hammouda, K., Kamel, M. (2006). Data Mining in e-learning. E-Learning Networked Environments and Architectures: A Knowledge Processing Perspective, Samuel Pierre (Ed.), Springer Book Series: Advanced Information and Knowledge Processing, 1-28.
[110]Hardof-jaffe, S., Hershkovitz, A., Abu-kishk, H., Bergman, O., Nachmias, R. (2009). How do students organize personal information

spaces?. In International Conference on Educational Data Mining, Cordoba, Spain, 250-258.

[111]Heathcote, E., Prakash, S. (2007). What your learning management system is telling you about supporting your teachers: monitoring system information to improve support for teachers using educational technologies at Queensland University Of Technology. In International Conference on Information Communication Technologies in Education, Samos Island, Greece, 1-6.

[112]Heathcote, E., Dawson, S. (2005). Data Mining for Evaluation, Benchmarking and Reflective Practice in a LMS. In World conference on E-learning in corporate, government, healthcare & higher education, Vancouver, Canada, 326-333.

[113]Heiner, C., Baker, R., Yacef, K. (2006). Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems, ITS2006,. Jhongli, Taiwan.

[114]Heiner, C., Heffernan, N., Barnes, T. (2007). Educational Data Mining Workshop at the 13th International Conference on Artificial Intelligence in Education, Los Angeles, California.

[115]Herlocker, J., Konstan, J., Tervin, L.G., Riedl, J. (2004). Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems Journal, 22, 1, 5-53.

[116]Heraud, J.M., France, L., Mille, A. (2004). Pixed: An ITS that guides students with the help of learners' interaction log. In International Conference on Intelligent Tutoring Systems, Maceio, Brazil, 57-64.

[117]Hershkovitz, A., Nachmias, R. (2008). Developing a log-based motivation measuring tool. In 1st International Conference on Educational Data Mining, Montreal, 226-233.

[118]Hershkovitz, A., Nachmias, R. (2009). Consistency of students' pace in online learning. In International Conference on Educational Data Mining, Cordoba, Spain, 71-80.

[119]Hien, N.T.N., Haddawy, P. (2007). A decision support system for evaluating international student applications. In Frontiers In Education Conference, Milwaukee, 1-6.

[120]Hsia, t., Shie, A., Chen, L. (2008). Course planning of extension education to meet market demand by using data mining techniques - an example of Chinkuo technology university in Taiwan. In Expert System Application Journal, 34,1, 596-602.

[121]Huang, C., Tsai, P., Hsu, C., Pan, R. (2006). Exploring cognitive difference in instructional outcomes using text mining technology. In IEEE conference on System, Man and Cybernertics, Taipei, Taiwan, 2116-2120.

[122]Huang, J., Zhu, A., Luo, Q. (2007). Personality mining method in web based education system using data minig. In IEEE International Conference on Grey Systems and Intelligent Services, Nanjing, China, 155-158.

[123]Huang, C., Lin, W., Wang, W., Wang, W. (2009). Planning of educational training courses by data mining: Using China Motor Corporation as an example. In Expert System with Application Journal, 36(3), 7199-7209.

[124]Huang, T., Cheng, S., Huang, Y. (2009b). A blog article recommendation generating mechanism using an SBACPSO algorithm. In Expert System with. Application Journal, 36(7), 10388-10396.

[125]Hübscher, R., Puntambekar, S., Nye, A. (2007). Domain Specific Interactive Data Mining. In Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling, Corfu, Greece, 81-90.

[126]Hurley, T., Weibelzahl, S. (2007). Using MotSaRT to support on-line teachers in student motivation. In European Conference on Technology-Enhanced Learning, Crete, Creece, 101-111.

[127]Hwang, W.Y., Chang, C.B., Chen, G.J. (2004). The relationship of learning traits, motivation and performance-learning response dynamics. In Computer & Education Journal, 42, 267-287.

[128]Hwang, G.J. (2005). A data mining approach to diagnosing student learning problems in science courses. In Journal of Distance Education Technologies, 3,4, 35-50.

[129]Hwang, G.J., Tsai, P.S., Tsai, C.C., Tseng, J.C.R. (2008). A novel approach for assisting teachers in analyzing student web-searching behaviors. In Computer & Education Journal, 51, 926-938.

[130]Ibrahim, Z., Rusli, D. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In Annual SAS Malaysia Forum, Kuala Lumpur, 1-6.

[131]Iksal S., Choquet C. (2005). Usage Analysis Driven by Models in a Pedagogical Context. Workshop on Usage Analysis in Learning Systems, In 12th International Conference on Artificial Intelligence, Amsterdam, 1- 8.

[132]Ingram, A. (1999). Using web server logs in evaluating instructional web sites. In Journal of Educational Technology Systems, 28,2, 137–157.

[133]Jin, H., Wu, T., Liu, Z., Yan, J. (2009). Application of Visual Data Mining in Higher-Education Evaluation System. In International Workshop on Education Technology and Computer Science, Washington, DC, 101-104.

[134]Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M. (2007). LOCO-Analyist: A tool for raising teacher's awareness in online learning environments. In European Conference on Technology-Enhanced Learning, Crete, 112-126.

[135]Jong, B.S., Chan, T.Y., Wu, Y.L. (2007). Learning log explorer in e-learning diagnosis. In IEEE Transactions on Education Journal, 50,3, 216-228.

[136]Jonsson, A., Hasmik, J. Johns, Mehranian, H., Arroyo, I., Woolf, B., Barto, A., Fisher, D., Mahadevan, S. (2005). Evaluating the Feasibility of Learning Student Models from Data, In Educational Data Mining AAAI Workshop, Pittsburgh, 1-6.

[137]Juan, A., Daradoumis, T., Faulin, J., Xhafa, F. (2009). SAMOS: a model for monitoring students' and groups' activities in collaborative e-learning. In International Journal of Learning Technology, 4,1-2, 53-72.

[138]Karampiperis, P., Sampson, D. (2005). Adaptive learning resources sequencing in educational hypermedia systems. In Educational Technology & Society Journal, 8,4, 128-147.

[139]Kay, J., Maisonneuve, N., Yacef, K., Zaiane, O.R. (2006). Mining Patterns of Events in Students' Teamwork Data. In Proceedings of Educational Data Mining Workshop. Taiwan, 1-8.

[140]Kelly, D., Tangney, B. (2005). First Aid for You: Getting to Know Your Learning Style Using Machine Learning. In IEEE international Conference on Advanced Learning Technologies, Washington, DC, 1-3.

[141]Khajuria, S. (2007). A Model to Predict Student Matriculation from Admissions Data. Master of Science, Ohio University, Industrial and Manufacturing Systems Engineering.

[142]Kiang, M.Y., Fisher, D.M., Chen, J.V., Fisher, S. A., Chi, R. T. (2009). The application of SOM as a decision support tool to identify AACSB peer schools. Decision Support System Journal, 47,1, 51-59.

[143]Kim, J., Chern, G., Feng, D., Shaw, E., Hovy, E. (2006). Mining and Assessing Discussions on the Web through Speech Act Analysis. In AAAI Workshop on Web Content Mining with Human Language Technologies, Athens, GA, 1-8.

[144]Kiu, C.C., Lee, C.S. (2007). Learning Objects Reusability and Retrieval through Ontological Sharing: A Hybrid Unsupervised Data Mining Approach. In IEEE Conference on Advanced Learning Technologies, Japan, 548-550.

[145]Koedinger, K., Cunningham, K., Skogsholm A., Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In 1st International Conference on Educational Data Mining, Montreal, 157-166.

[146]Kosba, E.M., Dimitrova, V., Boyle, R. (2005). Using Student and Group Models to Support Teachers in Web-Based Distance Education. In International Conference on User Modeling, Edinburgh, 124-133.

[147]Kotsiantis, S., Pierrakeas, C., Pintelas, P. (2003). Preventing student dropout in distance learning systems using machine learning techniques, In International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Oxford, 3-5.

[148]Kotsiantis, S.B., Pintelas, P.E., (2005). Predicting Students' Marks in Hellenic Open University. In IEEE international Conference on Advanced Learning Technologies, Washington, DC, 664-668.

[149]Khribi, M. K., Jemni, M., Nasraoui, O. (2008). Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In IEEE International

Conference on Advanced Learning Technologies, Washington, DC, 241-245.
[150]Kristofic, A. (2005). Recommender system for adaptive hypermedia applications. In Student Research Conference in Informatics and Information Technologies, Bratislava, 229-234.
[151]Lau, R., Chung, A., Song, D., Huang, Q. (2007). Towards fuzzy domain ontology based concept map generation for e-learning. In International Conference on Web-based Learning, Ediburgh, 90-101.
[152]Lee, C.S. (2007). Diagnostic, predictive and compositional modeling with data mining in integrated learning environments. In Computer & Education Journal, 49, 562-580.
[153]Lee, M.W., Chen, S.Y., Liu, X. (2007). Mining learners' behavior in accessing web-based interface. In International Conference Edutainment, Hong Kong, China, 226-346.
[154]Lee, C.H., Lee, G., Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. In Expert Systems Application Journal, 36, 1675-1684.
[155]Lee, M.W., Chen, S.Y., Chrysostomou, K., Liu, X. (2009b). Mining student's behavior in web-based learning programs. In Expert System with Applications Journal, 36, 3459-3464.
[156]Lemire, D., Boley, H. Mcgrath, S., Ball, M. (2005). Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. In International Journal of Interactive Technology and Smart Education, 2(3), 1-11.
[157]Licchelli, O., Basile, T. M., Di Mauro, N., Esposito, F., Semeraro, G., Ferilli, S. (2004). Machine learning approaches for inducing student models. In International Conference on innovations in Applied Artificial intelligence, Ottawa, Canada 935-944.
[158]Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. In Computer & Education Journal, 53,3, 950-965.
[159]Li, X., Luo, Q., and Yuan, J. 2007. Personalized Recommendation Service System in E-Learning Using Web Intelligence. In Proceedings of the 7th international Conference on Computational Science,Beijing, China, 531-538.
[160]Lin, F., Hsieh, L., Chuang, F. (2009). Discovering genres of online discussion threads via text mining. In Computer & Education Journal, 52,2, 481-495.
[161]Liu, F., Shih, B. (2007). Learning activity-based e-learning material recommendation system. In International Symposium on Multimedia, Taichung, Taiwan, 343-348.
[162]Lu, J. (2004). A personalized e-learning material recommender system. In International Conference on Information Technology for Application, China, 374-379.
[163]Lu, F., Li, X., Liu, Q., Yang, Z., Tan, G,. He, T. (2007). Research on Personalized E-Learning System Using Fuzzy Set Based Clustering Algorithm. In International Conference on Computational Science, Beijing, China, 587-590.
[164]Luan, J., (2002). Data mining, knowledge management in higher education, potential applications. In Workshop Associate of Institutional Research International Conference. Toronto, 1–18.
[165]Ma, Y., Liu, B., Wong, C., Yu, P. Lee, S. (2000). Targeting the right students using data mining. In KDD '00: Proceedings of the sixth ACM SIGKDD International Conference on Knowledge discovery and data mining, 457–464.
[166]Machado, L., Becker, K. (2003). Distance Education: A Web Usage Mining Case Study for the Evaluation of Learning Sites. In International Conference on Advanced Learning Technologies, Athens, Greece. 360-361.
[167] Madhyastha, T., Hunt, E. (2009). Mining Diagnostic Assessment Data for Concept Similarity. Journal of Educational Data Mining, 1, 1, 72-91.
[168]Markellou, P., Mousourouli, I., Spiros, S., Tsakalidis, A. (2005). Using semantic web mining technologies for personalized e-learning experiences. In Proceedings of the web-based education, Grindelwald, Switzerland, 461–826.
[169]Martinez, D., (2001). Predicting Student Outcomes Using Discriminant Function Analysis. In Meeting of the Research and Planning Group, Lake Arrowhead, CA, 1-22.
[170]Matsuda, N., Cohen, W., Sewall, J., Lacerda, G., Koedinger, K. R. (2007). Predicting students performance with SimStudent that learns cognitive skills from observation. In International Conference on Artificial Intelligence in Education, Amsterdam, Netherlands, 467-476.
[171]Mavrikis, M. (2008). Data-Driven Prediction of the Necessity of Help Requests in ILEs. In International Conference on Adaptive Hypermedia, Hannover, Germany, 316-319.
[172]Mazza, R., Vania, D. (2003). The design of a Course Data Visualizator: an Empirical Study. In International Conference on New Educational Environments, Lucerne, 215-220.
[173]Mazza, R., Milani, C. (2004). GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems, In International Conference on Technology Enhanced Learning, Milan, 1-8.
[174]Mazza, R. (2009). Introduction to Information Visualization. Springer.
[175]Mcdonald, B., (2004). Predicting student success. In Journal for Mathematics Teaching and Learning,1-14.
[176]Mclaren, B.M., Koedinger, K.R., Schneider, M., Harrer, A., Lollen, L. (2004). Bootstrapping Novice Data: Semi-Automated Tutor Authoring Using Student Log Files. In Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, Alagoas, Brazil, 1-10.
[177]Merceron, A, Oliveira, C., Scholl, M., Ullrich, C. (2004). Mining for Content Re-Use and Exchange - Solutions and Problems. In International Conference on Semantic Web Conference, Hiroshima, Japan, 1-2.
[178]Merceron, A., Yacef, K. (2004). Mining student data captured from a web-based tutoring tool: Initial exploration and results. Journal of Interactive Learning Research, 15,4, 319–346.
[179]Merceron, A., Yacef, K. (2005). Educational Data Mining: a Case Study. In International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands, 1-8.
[180]Merceron, A., Yacef, K. (2008). Interestingness Measures for Association Rules in Educational Data. In International Conference on Educational Data Mining, Montreal, Canada, 57-66.
[181]Mihaescu, C.,Burdescu, D. (2006). Testing attribute selection algorithms for classification performance on real data. In International IEEE conference Intelligent Systems, Varna, Bulgaria, 581-586.
[182]Minaei-bidgoli, B., Kashy, D.A., Kortmeyer, G., Punch, W.F. (2003). Predicting student performance: an application of data mining methods with an educational Web-based system. In International Conference on Frontiers in Education, 13-18.
[183]Minaei-bidgoli, B., Tan, P., Punch, W. (2004). Mining interesting contrast rules for a web-based educational system. In International Conference on Machine Learning Applications, Los Angeles, USA, 1-8.
[184]Myszkowski, P.B.; Kwasnicka, H., Markowska-kaczmar, U. (2008). Data Mining Techniques in e-Learning CelGrid System. In International Conference on Computer Information Systems and Industrial Management Applications. Ostrava, The Czech Republic. 315-319.
[185]Monk, D. (2005). Using data mining for e-learning decision making. Electronic Journal of e-Learning, 3,1, 41–54.
[186]Mor e., Minguillón J. (2004). E-learning Personalization based on Itineraries and Long-term Navigational Behavior. In Thirteenth World Wide Web Conference, New York, 264-265.
[187]Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell! In Proceedings of the Workshop on Educational Data Mining. 15–22.
[188]Mostow, J., Beck., J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. In Journal Natural Language Engineering, 12 ,2, 195-208.
[189]Muehlenbrok, M. (2005). Automatic action analysis in an interactive learning environment. In workshop on Usage Analysis in Learning Systems at the 12th International Conference on Artificial Intelligence in Education, Amsterdam The Netherlands, 73-80.
[190]Myller, N., Suhonen, J., Sutinen, E. (2002). Using Data Mining for Improving Web-Based Course Design. In International Conference on Computers in Education, Washington, 959- 964.

[191] Nagata, R. Takeda, K., Suda, K., Kakegawa, J., Morihiro, K. (2009). Edu-mining for book recommendation for pupils. In International Conference on Educational Data Mining, Cordoba, Spain, 91-100.
[192] Nankani, E., Simoff, S. Denize, S. Young, L. (2009). Supporting Strategic Decision Making in an Enterprise University Through Detecting Patterns of Academic Collaboration. In International United Information Systems Conference, Sydney, Australia, 496-507.
[193] Nebot, A., Castro, F., Vellido, A., Mugica, F. (2006). Identification of fuzzy models to predict students perfornance in an e-learning environment. In International Conference on Web-based Education, Puerto Vallarta, 74-79.
[194] Nesbit, J. C., Xu, Y., Winne, P. H., Zhou, M. (2008). Sequential pattern analysis software for educational event data. In International Conference on Methods and Techniques in Behavioral Research, Netherlands, 1-5.
[195] Novak, J.D., Cañas, A.J. (2006). The theory underlying concept maps and how to construct and use them. Technical Report IHMC CmapTools 2006-01.
[196] Nugent, R., Ayers, E., Dean, N. (2009). Conditional subspace clustering of skill mastery: identifying skills that separate students. In International Conference on Educational Data Mining, Cordoba, Spain, 101-110.
[197] Ogor, E.N. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. In Electronics, Robotics and Automotive Mechanics Conference, Washington, DC, 354-359.
[198] Oladokun, V.O., Adebanjo, A.T., Charles-owaba, O.E. (2008). Predicting student's academic performance using artificial neural network: A case study of an engineering course. In Pacific Journal of Science and Technology, 9,1, 72-79.
[199] Orzechowski, T., Ernst, S., Dziech, A. (2007). Profiled search methods for e-learning systems. In International Workshop on Learning Object Discovery & Exchange at European Conference on Technology Enhanced Learning, Crete, Greece, 1-10.
[200] Ouyang, Y., Zhu, M. (2007). eLORM: Learning Object Relationship Mining based Repository. In IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, Tokyo, Japan, 691-698.
[201] Pahl, C., Donnellan, C., (2003). Data mining technology for the evaluation of web-based teaching and learning systems. In Congress E-learning. Montreal, Canada, 1-7.
[202] Pardos, Z., Heffernan, N., Anderson, B., Heffernan, C. (2007). The Effect of Model Granularity on Student Performance Prediction Using Bayesian Networks. In International Conference on User Modeling, Corfu, Greece, 435-439.
[203] Pardos, Z., Beck, J.E., Ruiz, C., Heffernan, N. (2008). The Composition Effect: Conjunctive or Compensatory? An Analysis of Multi-Skill Math Questions in ITS. In International Conference on Educational Data Mining, Montreal, 147-156.
[204] Pardos, Z., Beck. J.E., Heffernan, N. (2009). Determining the significance of item order in randomized problem sets. In International Conference on Educational Data Mining, Cordoba, Spain, 111-120.
[205] Pavlik, P., Cen, H., Koedinger, K. (2009). Learning factors transfer analysis: using learning curve analysis to automatically generate domain models. In International Conference on Educational Data Mining, Cordoba, Spain, 121-130.
[206] Pechenizkiy, M., Calders, T., Vasilyeva, E., De bra, P. (2008). Mining the Student Assessment Data: Lessons Drawn from a Small Scale Case Study. In International Conference on Educational Data Mining, Montreal, 187-191.
[207] Pechenizkiy, M., Trcka, N., Vasilyeva, E., Aalst, W., De bra, P. (2009). Process mining online assessment data. In International Conference on Educational Data Mining, Cordoba, Spain, 279-288.
[208] Prata, D., Baker, R., Costa, E., Rose, C., Cui, Y. (2009). Detecting and understanding the impact of cognitive and interpersonal conflict in computer supported collaborative learning environments. In International Conference on Educational Data Mining, Cordoba, Spain, 131-140.
[209] Pritchard, D., Warnakulasooriya, R. (2005). Data from a Web-based Homework Tutor can predict Student's Final Exam Score. In World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, 2523-2529.
[210] Psaromiligkos, Y., Orfanidou, M. Kytagias, C., Zafiri, E. (2009). Mining log data for the analysis of learners' behaviour in web-based learning management systems. In Operational Research Journal, 1-14
[211] Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaïane, O. R. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. In IEEE Transaction on Knowledge and Data Engineering, 21,6, 759-772.
[212] Quevedo, J.R., Motañes, E. (2009). Obtaining rubric weitghts for assessments by more than one lecturer using a pairwise learning model. In International Conference on Educational Data Mining, Cordoba, Spain, 289-298.
[213] Raghavan, S.N.R. (2005). Data Mining in E-commerce: A Survey. Sadhana Journal, 30, 2&3, 275–289.
[214] Rahkila, M., Karjalainen, M. (1999). Evaluation of learning in computer based education using log systems. In ASEE/IEEE frontiers in education conference, San Juan, Puerto Rico, 16–21.
[215] Rai, D., Gong, Y., Beck, J.E. (2009). Using Dirichlet priors to improve model parameter plausibility. In International Conference on Educational Data Mining, Cordoba, Spain, 141-150.
[216] Ramli, A.A. (2005). Web usage mining using apriori algorithm: UUM learning care portal case. In International conference on knowledge management, Malaysia, 1-19.
[217] Ranjan, J., Khalil, S. (2008). Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective. In Information Technology Journal, 7,1, 16-23.
[218] Rallo, R. Gisbert, M., Salinas, J. (2005). Using data mining and social networks to analyze the structure and content of educative online communities. In International conference on multimedia and ICTs in education, Caceres, Spain, 1-10.
[219] Ritter, S., Harris, T., Nixon, T., Dickison, D., Murray, R., Towle, B. (2009). Reducing the knowledge tracing space. In International Conference on Educational Data Mining, Cordoba, Spain, 151-160.
[220] Robinet, V., Bisson, G., Gordon, M., Lemaire, B. (2007). Searching For Student Intermediate Mental Steps. In Workshop on Data Mining for User Modeling, at the International Conference on User Modeling Corfu, Greece, 101-105.
[221] Romero, C., Ventura, S., De Bra, P. (2004). Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction: The Journal of Personalization Research 14,5, 425–464.
[222] Romero, C., Ventura, S. (2006). Data mining in e-learning, Wit Press.
[223] Romero, C., Ventura, S. (2007). Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with Applications, 1, 33, 135-146.
[224] Romero, C., Pechenizkiy, M., Calders, T., Viola, S.R. (2007). International Workshop on Applying Data Mining in e-Learning (ADML'07) as part of the 2nd European Conference on Technology Enhanced Learning, EC-TEL2007, Crete, Creece.
[225] Romero, C., Ventura, S., Salcines, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computer & Education, 51(1), 368-384.
[226] Romero, c., ventura, s., hervás, c., gonzales, p. (2008). Data mining algorithms to classify students. In International Conference on Educational Data Mining, Montreal, Canada, 8-17.
[227] Romero, C., Gutierrez, S., Freire, M., Ventura, S. (2008). Mining and visualizing visited trails in web-based educational systems. In International Conference on Educational Data Mining, Montreal, Canada, 182-185.
[228] Romero, C., Gonzalez, P., Ventura, S., del Jesus, M.J., Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. In Expert System with Application Journal, 36, 1632-1644.
[229] Romero, C., Ventura, S., Zafra, A., de bra, P. (2009). Applying Web Usage Mining for Personalizing Hyperlinks in Web-based Adaptive Educational Systems. Computers & Education, 53, 3, 828-840.
[230] Romero, C. Ventura, S., Pechenizkiy, M., Baker, R. (2010). Handbook of Educational Data Mining. Taylor & Francis.
[231] Romesburg, H.C. (2004). Cluster Analysis for Researchers. Krieger Pub.

[232] Rosta, F., Brusilovsky, P. (2006). Social Navigation Support in a Course Recommendation System. In International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Dublin, 1-10.

[233] Reffay, C., Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In International Conference on Computer Supported Collaborative Learning, Bergen, Norvège, 1-6.

[234] Retalis, S., Papasalouros, A., Psaromilogkos, Y., Siscos, S., Kargidis, T. (2006). Towards networked learning analytics – A concept and a tool. In Fifth International Conference on Networked Learning, 1-8.

[235] Reyes, P., Tchounikine, P. (2005). Mining learning groups' activities in forum-type tools. In Conference on Computer Support For Collaborative Learning: Learning, Taipei, Taiwan, 509-513.

[236] Rus, V., Lintean, M., Azevedo, R. (2009). Automatic detection of student mental models during prior knowledge activation in MetaTutor. In International Conference on Educational Data Mining, Cordoba, Spain, 161- 170.

[237] Saini, P.S., Sona, D., Veeramachaneni, S., Ronchetti, M. (2005). Making E-Learning Better Through Machine Learning. In International Conference on Methods and Technologies for Learning, Barcelona, Spain, 1-6.

[238] Sanjeev, A.P., Zytkow, J.M. (1995). Discovering enrollment knowledge in university databases. In International Conference on Knowledge Discovery and Data Mining, Montreal, Canada, 246-251.

[239] Schönbrunn, K., Hilbert, A. (2007). Data Mining in Higher Education. In Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, 489-496.

[240] Schoonenboom, J., Heller, K, Neenoy, M., Levene, M. (2007). Trails in education: technologies that support navigational learning, Rotterdam, Sense Publisher.

[241] Selmoune, N., Alimazighi, Z. (2008). A decisional tool for quality improvement in higher education. In International Conference on Information and Communication Technologies, Damascus, Syria 1-6.

[242] Shangping, D., Ping, Z. (2008). A data mining algorithm in distance learning. In International Conference on Computer Supported Cooperative Work in Design, Xian, 1014-1017.

[243] Sheard, J., Ceddia, J., Hurst, J., Tuovinen J. (2003). Inferring Student Learning Behaviour from Website Interactions: A Usage Analysis. In Journal Education and Information Technologies, 8,3, 245-266.

[244] Shen, R., Yang, F., Han, P. (2002). Data analysis center based on e-learning platform. In Workshop The Internet Challenge: Technology and Applications, Berlin, Germany, 19–28.

[245] Shen, R., Han, P., Yang, F., Yang, Q., Huang, J. (2003). Data mining and case-based reasoning for distance learning. Journal of Distance Education Technologies 1,3, 46–58.

[246] Shen, L., Shen, R. (2004). Learning content recommendation service based-on simple sequencing specification. In International Conference on Web-based Learning, Beijing, China, 363-370.

[247] Simko, M., Bielikova, M. (2009). Automatic concept telationships discovery for an adaptive e-course. In International Conference on Educational Data Mining, Cordoba, Spain, 171-178.

[248] Singley, M.K., Lam, R.B. (2005). The Classroom Sentinel: supporting data-driven decision-making in the classroom. In 13th World Wide Web Conference, Chiba, Japan, 315-322.

[249] Song, D., Lin, H., Yang, Z. (2007). Opinion Mining in e-Learning System. In International Conference on Network and Parallel Computing Workshops, Washington, DC, 788-792.

[250] Spacco, J., Winters, T., Payne, T. (2006). Inferring use cases from unit testing. In Workshop on Educational Data Mining, New York, 1-7.

[251] Stamper, J., Barnes, T. (2009). Unsupervised MDP value selection for automating ITS capabilities. In International Conference on Educational Data Mining, Cordoba, Spain, 180-188.

[252] Stevens, R., Giordani, A., Cooper, M., Soller, A., Gerosa, L., Cox, C. (2005). Developing a Framework for Integrating Prior Problem Solving and Knowledge Sharing Histories of a Group to Predict Future Group Performance. In International Conference on Collaborative Computing: Networking, Applications and Worksharing, Boston, 1-9.

[253] Su, Z., Song, W., Lin, M., Li, J. (2008). Web Text Clustering for Personalized E-learning Based on Maximal Frequent Itemsets. In International Conference on Computer Science and Software Engineering, Washington, DC, 452-455.

[254] Superby, J.F., Vandamme, J.P., Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In International conference on intelligent tutoring systems, Educational Data Mining Workshop, Taiwan, 1-8.

[255] Tai, D.W., Wu, H.J., Li, P.H. (2008). Effective e-learning recommendation system based on self-organizing maps and association mining. In The Electronic Library Journal, 26,3, 329-344-

[256] Talavera, L., Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In Workshop on Artificial Intelligence in CSCL, Valencia, Spain, 17–23.

[257] Tang, C., Lau, R.W.H., Li., Q., Yin, H., Li, T., Kilis, D. (2000). Personalized courseware construction based on web data mining. In First International Conference on Web Information Systems Engineering, Hong Kong, China, 204-211.

[258] Tang, T. Y., Mccalla, G. (2002). Student modeling for a web-based learning environment: a data mining approach. In Conference on Artificial intelligence, Edmonton, Canada, 967-968.

[259] Tang, T., Mccalla, G. (2005). Smart recommendation for an evolving e-learning system. In International Journal on E-Learning 4,1, 105–129.

[260] Thomas, E.H., Galambos, N. (2004). What satisfies students? Mining student-opinion data with regression and decision tree analysis. In Research in Higher Education Journal, 45,3, 251-269.

[261] Tian, F., Wang, S., Zheng, C., Zheng, Q. (2008). Research on e-learning personality group based on fuzzy clustering analysis. In International Conference on Computer Supported Cooperative Work in Design, China, 1035- 1040.

[262] Tsai, C.J., Tseng, S.S., Lin, C.Y. (2001). A Two-phase fuzzy mining and learning algorithm for adaptive learning environment. In International Conference on Computational Science, San Francisco, 429-438.

[263] Tsantis, L., Castellani, J. (2001). Enhancing learning environments through solution-based knowledge discovery tools. In Journal of Special Education Technology, 16,4, 39-52.

[264] Tseng, S.S., Sue, P.C., Su, J.M., Weng, J.F., Tsai, W.N. (2007). A new approach for constructing the concept map. In Computers & Education Journal, 49, 691-707.

[265] Ueno, M., Nagaoka, K. (2002). Learning log database and data mining system for e-learning – on line statistical outlier detection of irregular learning processes. In International Conference on Advanced Learning Technologies, Tatarstan, Russia 436-438.

[266] Ueno, M. (2004). Data Mining and Text Mining Technologies for Collaborative Learning in an ILMS "Samurai". In IEEE international Conference on Advanced Learning Technologies, Washington, DC, 1052-1053.

[267] Vee, M.N., Meyer, b., Mannock, K.L. (2006). Understanding novice errors and error paths in Object-oriented programming through log analysis. In Workshop on Educational Data Mining, Taiwan, 13-20.

[268] Vellido, A., Castro, F., Etchells, T.A., Nebot, a., Mugica, F. (2007). Data mining of virtual campus data. In Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence (SCI) 62, series Advanced Information and Knowledge Processing. Springer. 223-254.

[269] Ventura, S., Romero, C., Hervas, C. (2008). Analyzing rule evaluation measures with educational datasets: a framework to help the teacher. In International Conference on Educational Data Mining, Montreal, Canada, 177-181.

[270] Vialardi, C., Bravo, J., Ortigosa, A. (2008). Improving AEH courses through log analysis. In Journal of Universal Computers Science, 14 (17), 2777-1798.

[271] Vialardi, C., Bravo, J., Shafti, L., Ortigosa, A. (2009). Recommendation in higher education using data mining techniques. In International Conference on Educational Conference, Cordoba, Spain, 190-198.

[272] Viola, S. R., Graf, S., Kinshuk, and Leo, T. 2006. Analysis of Felder-Silverman Index of Learning Styles by a Data-Driven Statistical Approach. In Proceedings of the Eighth IEEE international Symposium on Multimedia, Washington, DC, 959-964.

[273]Vranic, M., Pintar, D., Skocir, Z. (2007). The use of data mining in education environment. In International Conference on Telecommunications. Zagred, 243-250.
[274]Wang, F.H. (2002). On Using Data-Mining Technology for Browsing Log File Analysis in Asynchronous Learning Environment. In World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, 2005-2006.
[275]Wang., F.H. (2004). A fuzzy neural network for item sequencing in personalized cognitive scaffolding with adaptive formative assessment. In Expert System with Application Journal, 27, 11-25.
[276]Wang., F.H. (2008). Content recommendation based on education-contextualized browsing events for web-based personalized learning. In Educational Technology & Society Journal, 11,4, 94-112.
[277]Wang, A.Y., Newlin, M.H. (2002). Predictors of we-based performance: the role olf self-effecacy and reasons for taking an on-line class. In Computers in Human Behavior Journal, 18, 151-163.
[278]Wang, F.H., Shao, H.M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. In Expert System in Application Journal, 27, 365-377.
[279]Wang, W., Weng, J., Su, J., Tseng, S. (2004). Learning portfolio analysis and mining in scorm compliant environment. In ASEE/IEEE Frontiers in Education Conference, Savannah, Georgia, 17–24.
[280]Wang, y., Cheng, y., Chang, T., Jen, S.M. (2008). On the application of data mining technique and genetic algorithm to an automatic course scheduling system. In IEEE Conference on Cybernetics and Intelligent Systems, Chengdu, 400-405.
[281]Wang, Y., Tseng, M., Liao, H. (2009). Data mining for adaptive learning sequence in English language instruction. In Expert Systems with Applications Journal, 36, 7681-7686.
[282]Want, T., Mitrovic, A. (2002). Using Neural Networks to Predict Student's Performance. In International Conference on Computers in Education, Washington, DC, 1-5.
[283]Winters, T., Shelton, C. R., Payne, T., Mei, G. (2005). Topic extraction from item-level grades. In AAAI-05 Workshop on Educational Data Mining, Pittsburgh, 7-14.
[284]Wu, A., Leung, C. (2002). Evaluating learning behavior of Web-Based Training (WBT) using Web log. In International Conference on Computers in Education, New Zealand, 736- 737.
[285]Yamanishi, K., Li, H. (2001). Mining from open answers in questionnaire data. In Proceedings of the Seventh ACM SIGKDD international Conference on Knowledge Discovery and Data Mining , San Francisco, California, 443-449.
[286]Yang, T. D., Lin, T., Wu, K., (2002). An Agent-Based Recommender System for Lesson Plan Sequencing. In International Conference on Advanced Learning Technologies, Kazan, Russia, 14-20.
[287]Yang, F., Han, P., Shen, R., Hu, Z. (2005). A novel resource recommendation system based on connecting to similar e-learners. In International Conference on Web-based Learning, Hong Kong, China, 122-130.
[288]Yoo, J., Yoo, S., Lance, C., Hankins, J. (2006). Student progress monitoring tool using treeview. In Technical Symposium on Computer Science Education, ACM-SIGCESE, 373 - 377.
[289]Yudelson, M.V., Medvedeva, O., Legowski, E., Castine, M., Jukic, D., Rebecca C. (2006). Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS. In AAAI Workshop on Educational Data Mining, Boston, 1-8.
[290]Yu, C.H., Jannasch-pennell, A., Digangi, S., Wasson, B. (1999). Using On-line interactive statistics for evaluating Web-based instruction, In Journal of Educational Media International, 35, 157-161.
[291]Yu, P., Own, C., Lin, L. (2001). On learning behavior analysis of web based interactive environment. In International Conference on Computer and Electrical Engineering, Oslo/Bergen, Norway, 1-9.
[292]Yu. C.H., Digangi, S., Jannasch-pennell, A.K., Kaprolet, C. (2008). Profiling students who take online courses using data mining methods. In Online Journal of Distance Learning Administration, XI(II), 1-14.
[293]Zafra, A., Ventura, S. (2009), Predicting student grades in learning management systems with multiple instance programming. In International Conference on Educational Data Mining, Cordoba, Spain, 307-314.
[294]Zaïane, o., Luo, J. (2001). Web usage mining for a better web-based learning environment. In Proceedings of Conference on Advanced Technology for Education. Banff, Alberta, 60–64.
[295]Zaïane, O. (2002). Building A Recommender Agent for e-Learning Systems. In Proceedings of the International Conference in Education, Auckland, New Zealand, 55-59.
[296]Zakrzewska, D. (2008). Cluster analysis for user's modeling in intelligent e-learning systems. In International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, Poland, 209-214.
[297]Zhang, K., Cui, L., Wang, H., Sui, Q. (2007). An improvement of matrix-based clustering method for grouping learners in e-learning. In International Conference on Computer Supported Cooperative Work Design, Melbourne, Australia, 1010-1015.
[298]Zhang, C., Zhang, S. (2002). Association rule mining: models and algorithms. Lecture Notes in Artificial Intelligence, Springer.
[299]Zhang, X., Mostow, J., Duke, N., Trotochaud, C., Valeri, J., Corbett, A. (2008). Mining Free-form Spoken Responses to Tutor Prompts. In International conference on Educational Data Mining, Montreal, 234-241.
[300]Zhang, L, Liu, X., Liu, X. (2008b). Personalized instructing recommendation system based on web mining. In International Conference for Young Computer Scientists, Hunan, China, 2517-2521.
[301]Zheng, S. Xiong, S., Huang, Y., Wu, S. (2008). Using methods of association rules mining optimization in web-based mobile learning system. In International Symposium on Electronic Commerce and Security. Guangzhou, China, 967- 970.
[302]Zhu, F, Ip, H., Fok, A., Cao, J. (2007). PeRES: A personalized recommendation education system based on multi-agents & SCORM. In International Conference on Web-based Learning, Ediburgh, 31-42.
[303]Zinn, C., Scheuer, O. (2006). Getting to know your students in distance-learning contexts. In 1st European Conference on Tehcnology Enhanced Learning, 437-451.
[304]Zoubek, L., Burda, M. (2009). Visualization of differences in data measuring mathematical skills. In International Conference on Education Data Mining, Cordoba, Spain, 315-324.
[305]Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., Segovia, J. (2005). Web usage mining project for improving web-based learning sites. In International Conference on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, 205-210.
[306]Zukhri, Z., Omar, K. (2007). Solving new student allocation problem with genetic algorithms: A hard problem for partition based approach. In Journal of Zhejiang University. (1-9.

Cristobal Romero received the B.Sc and Ph.D. degree in computer science from the University of Granada, Spain, in 1996 and 2003, respectively.

He is currently an Associate Professor in the Department of Computer Science and Numerical Analysis, University of Cordoba.

He is an assistant professor in the Computer Science Department of Cordoba University. He has published more than 30 papers about Educational Data Mining in international journals and conferences. His current research interest is focused on applying data mining in e-learning systems.

Dr. Romero is member of the IEEE Computer, the International EDM Working Group and the steering committee of the EDM Conferences.

Sebastian Ventura received the B.Sc. and Ph.D. degrees in sciences from the University of Córdoba, Spain, in 1989 and 1996, respectively.

He is currently an Associate Professor in the Department of Computer Science and Numerical Analysis, University of Cordoba, where he heads the Knowledge Discovery and Intelligent Systems Research Laboratory. He has more than 70 international publications, 20 of them published in international journals. He has also worked on eleven research projects (being the coordinator of two of them) supported by the Spanish and Andalusian governments and the European Union. His current main research interests are in the fields of soft-computing, machine learning, data mining and its applications.

Dr. Ventura is asenior member in the IEEE Computer, Computational Intelligence and Systems, Man and Cybernetics societies and the Association of Computing Machinery (ACM).