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

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

***Abstract*—Educational data mining (EDM) 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. First, 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*—Data mining (DM), educational data mining (EDM), educational systems, knowledge discovery.**

I. Introduction

EDUCATIONAL data mining (EDM) is a field that exploits statistical, machine-learning, and data-mining (DM) algorithms over the different types of educational data. Its main objective is to analyze these types of data in order to resolve educational research issues [27]. 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's information have created large repositories of data reflecting how students learn [143]. 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 [186]. 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, which are as follows.

1) 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's behavior/performance, curriculum, etc., that was gathered in classroom environments.
2) 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's data stored by these systems in log files and databases.
3) Intelligent tutoring system (ITS) and adaptive educational hypermedia system (AEHS) are an alternative to the just-put-it-on-the-web approach, trying to adapt teaching to the needs of each particular student. DM 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 DM, like business, genetics, medicine, etc., because it follows the same steps as the general DM process [219]: preprocessing, DM, and postprocessing. However, it is important to note that in this paper, the term DM is used in a larger sense than the original/traditional DM definition, i.e., 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 describe other approaches, such as regression, correlation, visualization, etc., which 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 are not necessarily seen universally as being DM [20].

From a practical point of view, EDM allows, for example, to discover new knowledge based on students' usage data in order to help to validate/evaluate educational systems, to potentially improve some aspects of the quality of education, and to lay the groundwork for a more effective learning process [219]. Some similar ideas were already successfully applied in e-commerce systems, the first and most popular application of DM [211], in order to determine clients' interests so as to be able to increase online sales. However, to date, there has been comparatively less progress in this direction in education, although this situation is changing and there is currently an increasing interest in applying DM to the educational environment [228]. Even so, there are

Manuscript received October 8, 2009; revised March 24, 2010 and June 2, 2010; accepted June 14, 2010. Date of publication July 26, 2010; date of current version October 15, 2010. This work was supported by the Spanish Department of Research under Project TIN2008-06681-C06-03 and Project P08-TIC-3720 and by the Fondo Europeo de Desarrollo Regional (FEDER) funds. This paper was recommended by Associate Editor R. Alhajj.

The authors are with the Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba 14071, Spain (e-mail: comero@uco.es; sventura@uco.es).

Digital Object Identifier 10.1109/TSMCC.2010.2053532

some important issues that differentiate the application of DM, specifically to education, from how it is applied in other domains [221].

1) *Objective:* The objective of DM in each application area is different. For example, in EDM, there are 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.
2) *Data:* In educational environments, there are many different types of data available for mining. These data are specific to the educational area, and therefore 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, which 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.
3) *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 DM 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 DM [104]. 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 [177]. 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.

Nowadays, there is a great variety of educational systems/ environments such as: the traditional classroom, e-learning, LMS, adaptive hypermedia (AH) educational systems, ITS, tests/quizzes, texts/contents, and others such as: learning object (LO) repositories, concept maps, social networks, forums, educational game environments, virtual environments, ubiquitous computing environments, etc. All data provided by each of the aforementioned educational environments are different, thus enabling different problems and tasks to be resolved using DM 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.

TABLE I
EDM USERS/STAKEHOLDERS

Users/Actors	Objectives for using data mining
Learners/ Students/ Pupils	To personalize e-learning; to recommend activities to learners 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, 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.).

On the other hand, the International Working Group in EDM (http://www.educationaldatamining.org) has achieved the establishment of an annual International Conference on EDM in 2008, EDM'08 [19], EDM'09 [27], and EDM'10 [22]. This conference has evolved from previous EDM workshops at the AIED'07 [112], the EC-TEL'07 [222], the ICALT'07 [35], the UM'07 [17], the AAII'06 [34], the ITS'06 [111], the AAAI'05 [33], the AIED'05 [62], the ITS'04 [32], and the ITS'00 [30] conferences.

The number of publications about EDM has grown exponentially in the past few years (see Fig. 1). A clear sign of this tendency is the appearance of the peer-reviewed *Journal of Educational Data Mining* (JEDM) and two specific books on EDM edited by Romero and Ventura entitled: *Data Mining in E-learning* [220] and *The Handbook of Educational Data Mining* [228] co-edited by Baker and Pechenizkiy. There were also two surveys carried out previously about EDM. The first one [221] is a former review of Romero and Ventura with 81 references until 2005 in which papers were classified by the DM techniques used. In fact, this survey is an improved, updated, and much extended version of this previous one with 306 references

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

Type of Data/ Environment	References
Traditional Education	[32], [42], [66], [68], [78], [94], [97], [102], [117], [118], [121], [128], [131], [139], [140], [145], [146], [162], [163], [167], [173], [195], [196], [210], [215], [236], [237], [239], [252], [258], [261], [269], [271], [278], [290], [304].
Web-based Education/ E-learning	[11], [45], [49], [50], [63], [64], [85], [91], [96], [99], [101], [103], [116], [120], [127], [130], [144], [147], [151], [153], [154], [155], [156], [157], [175], [179], [180], [181], [188], [191], [197], [199], [212], [214], [225], [238], [240], [246], [253], [259], [263], [272], [275], [276], [284], [285], [286], [288], [289], [292], [293], [295], [298], [300].
Learning Management Systems	[28], [46], [48], [59], [67], [76], [100], [109], [110], [132], [159], [164], [168], [171], [178], [182], [183], [208], [209], [223], [224], [232], [242], [254], [266], [267], [274], [291], [303].
Intelligent Tutoring Systems	[9], [15], [16], [18], [26], [29], [31], [47], [61], [65], [83], [98], [106], [114], [124], [134], [143], [174], [177], [185], [200], [203], [213], [217], [218], [234], [249], [265], [280], [287], [294].
Adaptive Educational Systems	[4], [23], [37], [38], [69], [92], [93], [105], [123], [125], [133], [136], [138], [148], [160], [161], [187], [219], [227], [245], [257], [260], [268], [277], [279], [301].
Tests/ Questionnaires/	[7], [12], [14], [25], [41], [43], [51], [54], [57], [79], [88], [126], [165], [194], [201], [202], [204], [205], [248], [270], [281], [283], [302].
Texts/ Contents	[1], [3], [40], [73], [107], [141], [150], [158], [235], [247], [251], [264], [283], [297].
Others	[2], [13], [44], [53], [55], [71], [74], [77], [108], [122], [137], [142], [152], [190], [198], [206], [216], [231], [233], [250], [262], [299].

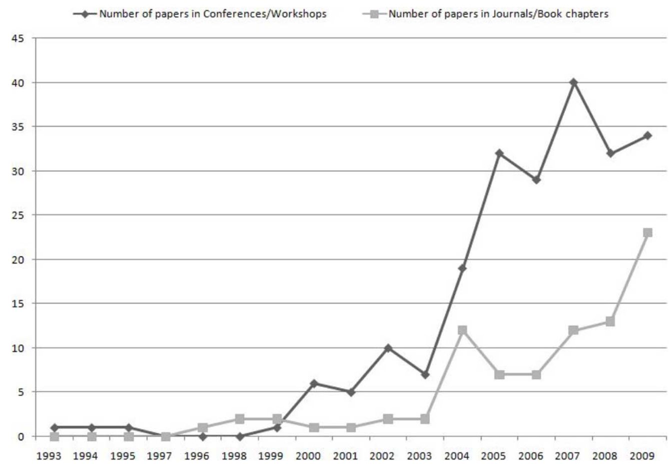

Fig. 1. Number of published papers until 2009 grouped according to the year. Note that we have counted only 300 papers in our reference section and not the total number of papers that were really published about EDM.

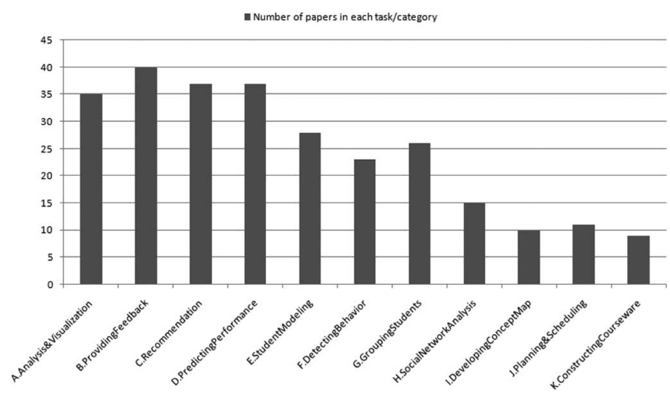

Fig. 2. Number of published papers until 2009 grouped by task/category. Note that we have counted only 300 papers in our reference section and not the total number of papers actually published about EDM.

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 Baker and Yacef with 46 references encompassing up to 2009. This survey uses mainly the top eight most cited papers in the first 2005 review and the Proceedings of the EDM'08 and the EDM'09 conferences; it also groups papers according to EDM methods and applications, as we describe in Section II.

Finally, it is important to highlight that most of the pioneer and older research (from 1993 to 1999) deals with predicting student's performance (see Task D in Section II). In fact, there is a huge body of studies on this topic in educational journals and conferences; and although seminal works date back to decades ago, new developments are highly relevant.

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

II. EDUCATIONAL TASKS AND DM 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, and 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 *et al.* [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 Fig. 2) for the main educational tasks that have employed DM techniques. These categories come from different research communities (as we have previously described in Section I), and they also use different DM tasks and techniques. On 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 sections 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.

As we can see in Fig. 2, the categories or research lines that have the most papers published are the first eight ones (from A to G with 23 or more references each), and the categories that have the fewest papers published are the last four (from H to K with less than 15 references). We think that this may be mainly due to the fact that the first eight categories are older than the last four (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 three. We also want to point out that we have organized these categories by grouping them near the most closely related ones, which 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 [86]. 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's behavior [282]. It is not surprising that teachers prefer pedagogically oriented statistics (overall success rate, mastery levels, typical misconceptions, percentage of exercises tackled, and material read) that are easy to interpret [301]. 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 things such as: where students enter and exit, the most popular pages, the browsers students tend to use, and patterns of use over time, [130]; the number of visits, origin of visitors, number of hits, and patterns of use throughout various time periods [95]; number of visits and duration per quarter, top search terms, and number of downloads of e-learning resources [99]; number of different pages browsed and total time for browsing different pages [127]; usage summaries and reports on weekly and monthly user trends and activities [183]; session statistics and session patterns [199]; statistical indicators on the learner's interactions in forums [5]; the amount of material students might go through and the order in which students study topics [212]; resources used by students and resources valued by students [241]; the overall averages of contributions to discussion forums, the amount of posting versus replies, and the amount of learner-to-learner interaction versus learner-to-teacher interaction [110]; the time a student dedicates to the course or a particular part of it [199]; the learners' behavior and time distribution and the distribution of network traffic over time [303]; and the frequency of studying events, patterns of studying activity, timing and sequencing of events, and the content analysis of students' notes and summaries [103]. Statistical analysis is also very useful to obtain reports assessing [81] 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 to understand and analyze data [172]. 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 [205]; mean values of attributes analyzed in data to measure mathematical skills [302]; tutor–student interaction data from an automated reading tutor [185]; statistical graphs about assignments complement, questions admitted, exam score, etc. [242]; student tracking data regarding social, cognitive, and behavioral aspects of students [170]; student's attendance, access to resources, overview of discussions, and results on assignments and quizzes [171]; weekly information regarding students' and groups' activity [135]; student's 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 [286] and higher education student-evaluation data [131]; student's interactions with online learning environments [132]; the students' online exercise work, including students' interactions and answers, mistakes, teachers' comments, etc. [176]; 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 [225]; and the sequence of LOs and educational trails [238].

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 [296].

There are many studies that apply/compare several DM 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 [292]. Association analysis, clustering analysis, and case-based reasoning have also been used to organize course material and assign homework at different levels of difficulty [243]. 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 [232]. 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's performance [215]. Several classifier algorithms have been applied to predict whether the teacher will recommend an intervention strategy for motivational profiles [124]. Clustering and association rules have been used in the academic community to potentially improve some qualitative teaching aspects [271].

Association-rule mining has been used to confront the problem of continuous feedback in the educational process [208]; to analyze learning data and to figure out whether students use resources and possibly whether their use has any (positive) impact on marks [178]; to determine the relationship between each learning-behavior pattern so that the teacher can promote collaborative learning behavior on the Web [289]; to find embedded information, which can be provided to teachers to further analyze, refine, or reorganize teaching materials and tests in adaptive learning environments [260]; to optimize the content of the university e-learning portal [214]; to discover interesting associations between student attributes, problem attributes, and solution strategies in order to improve online education systems for both teachers and students [181]; to analyze rule-evaluation measures in order to discover the most interesting rules [267]; to identify interesting and unexpected learning patterns, which in turn may provide decision lines, enabling teachers to more efficiently organize their teaching structure [272]; to provide feedback to the course author about how to improve courseware [219]; 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's web-usage patterns [46]; to improve an adaptive course design in order to show recommendations on how to enhance the course structure and contents [268]; to find interesting relationships between attributes, solution strategies adopted by learners, etc., from a web-based mobile learning system [299]; to help the teacher to discover beneficial or detrimental relationships between the use of web-based educational resources and student's learning [226]; to reveal information about university students' enrollment [236]; to help organizations to determine the thinking styles of learners and the effectiveness of a website structure [101]; to evaluate educational website design [164]; and to mine open answers in questionnaire data in order to analyze surveys [283].

Other different DM techniques have been applied to provide feedback such as: domain-specific interactive DM to find the relationships between log data and student's behavior in an educational hypermedia system [123]; temporal DM to describe, interpret, and predict student's 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 [84]; timely alerts to detect critical teaching and learning patterns and to help teachers to make sense of what is happening in their classrooms [246]; and usage data analysis to improve the effectiveness of the learning process in e-learning systems [182].

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's assessment data [204]. A group of DM 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 to understand the main factors influencing learner's 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 [248]. Hierarchical clustering has been used for mining multiple-choice assessment data for similarity of the concepts represented by the responses [165]. Common-factor analysis and collaborative filtering have been used to discover the fundamental topics of a course from item-level grades [281]. 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/DM to educational data is to analyze educational contents, to summarize/analyze the learner's 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 [247]; 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 [297]; to facilitate the automatic coding process of an online discussion forum [158]; for collaborative learning prompted by learners' comments on discussion boards [264]; 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 [298]; to study eye movements (of students' reading concept maps) in order to detect when focal actions overlap unrelated actions [192]; 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 [137]; to generate personalized activities for learners [277]; for personalizing based on itineraries and long-term navigational behavior [184]; to recommend the most appropriate future links for a student to visit in a web-based adaptive educational system [227]; to include the concept of recommended itinerary in Sharable Content Object Reference Model (SCORM) standard by combining teachers' expertise with learned experience [184]; to select different LOs for different learners based on learner's profiles and the internal relation of concepts [244]; for personalizing activity trees according to learning portfolios in a SCORM compliant environment [277]; for recommending lessons (LOs or concepts) that a student should study next while using an AH system [148]; to discover LO relationship patterns to recommend related LOs to learners [198]; and for adapting learning resource sequencing [136].

Association-rule mining has been used to recommend online learning activities or shortcuts on a course website [293]; to produce recommendations for learning material in e-learning systems [166]; for content recommendation based on educationally contextualized browsing events for web-based personalized learning [274]; 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 [92]; for building a personalized e-learning material-recommender system to help students to find learning materials [160]; for course recommendation with respect to optimal elective courses [253]; and for designing a material recommendation system based on the learning actions of previous learners [159].

Clustering has been developed to establish a recommendation model for students in similar situations in the future [276]; for grouping Web documents using clustering methods in order to personalize e-learning based on maximal frequent item sets [251]; for providing personalized course material recommendations based on learner's ability [161]; and to recommend to students those resources they have not yet visited, but would find most helpful [96].

Other DM techniques used are: neural networks and decision trees to provide adaptive and personalized learning support [100]; production rules to help students to make decisions about their academic itineraries [269]; decision tree analysis to recommend optimal learning sequences to facilitate the students' learning process and maximize their learning outcome [279]; learning factor transfers and Q-matrixes to generate domain models that will sequence item types to maximize learning [203]; an item-order effect model to suggest the most effective item sequences to facilitate learning [202]; 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 [284]; DM and text mining to recommend books related to the books that the target pupil has consulted [189]; case-based reasoning to offer contextual help to learners, providing them with an adapted link structure for the course [114]; 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) [249]; and an extended serial blog article composition particle swarm optimization (SBACPSO) algorithm to provide optimal recommended materials to users in blog-assisted learning [122].

D. Predicting Student's 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 ITSs [106]. Other comparisons of different DM algorithms are made to classify students (predict final marks) based on Moodle usage data [224]; to predict student's performance (final grade) based on features extracted from logged data [180]; and to predict university students' academic performance [128].

Different types of neural-network models have been used to predict final student grades (using back-propagation and feed-forward neural networks) [94]; to predict the number of errors a student will make (using feedforward and backpropagation) [280]; to predict performance from test scores (using backpropagation and counter propagation) [78]; 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) [196].

Bayesian networks have been used to predict student–applicant performance [102]; to model user knowledge and predict student's performance within a tutoring system [200]; to predict a future graduate's cumulative grade point average based on applicant background at the time of admission [117]; to model two different approaches to determine the probability a multiskill question has of being corrected [201] and to predict future group performance in face-to-face collaborative learning [250]; to predict end-of-year exam performance through student's activity with online tutors [12]; and to predict item response outcome [69].

Different types of rule-based systems have been applied to predict student's performance (mark prediction) in an e-learning environment (using fuzzy-association rules) [191]; to predict learner's performance based on the learning portfolios compiled (using key-formative assessment rules) [51]; for prediction, monitoring, and evaluation of student's academic performance (using rule induction) [195]; to predict final grades based on features extracted from logged data in an education web-based system (using genetic algorithm to find association rules) [240]; to predict student's grades in LMSs (using grammar-guided genetic programming) [291]; to predict student's performance and provide timely lessons in web-based e-learning systems (using decision tree) [45]; and 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) [146]; for predicting end-of-year accountability assessment scores (using linear regression prediction models) [7]; to predict student's performance from log and test scores in web-based instruction (using a multivariable regression model) [288]; for predicting student's academic performance (using stepwise linear regression) [97]; 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) [167]; for predicting university students' satisfaction (using regression and decision trees analysis) [258]; for predicting exam results in distance education courses (using linear regression) [188]; 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) [88]; 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) [79]; 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 online classes [275]; to predict a student's final exam score in online tutoring [207]; and for predicting high school students' probabilities of success in university [173].

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. DM 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 [89]. Different DM techniques and algorithms have been used for this task (mainly, Bayesian networks).

Several DM algorithms (naïve Bayes, Bayes net, support vector machines, logistic regression, and decision trees) have been compared to detect student mental models in ITSs [234]. 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 [115].

Bayesian networks have been used to make predictions about student's 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 [91]; to predict whether a student will answer a problem correctly [134]; 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 [98].

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 [218].

Association-rule algorithms have been applied for personality mining based on web-based education models in order to deduce learners' personality characteristics [120] and for student modeling in ITSs [168].

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) [83]. 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 [155]. The Markov decision process has been proposed to automatically create student models by generating hints for an IT that learns [26]. Fuzzy techniques have used student models in web-based learning environments in order to generate advice for the teachers [144]. A dynamic learning-response model has been developed for inferring, testing, and verifying student's learning models on an adaptive learning website [125]. Bootstrapping novice data can create an initial skeletal model of a tutor from log data collected from actual use of the tool by students [174]. A collaborative-based DM approach has been developed for diagnostic and predictive student modeling purposes in integrated learning environments [151]. Multiple correspondence analysis and cross validation by correlation analysis have been applied to identify learning styles in Index of Learning Styles (ILS) questionnaires [270]. The Q-matrix method has been used to create concept models that represent relationships between concepts and questions, and to group student's 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's knowledge [213]. Self-organizing maps and principal component analysis have been applied for predictive and compositional modeling of the student's profile [150]. A clustering algorithm (K-means) has been developed to model student's behavior with a very small set of parameters without compromising the behavior of the system [217].

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's behavior are decision tree neural networks, naïve Bayes, instance-based learning, logistic regression, and support vector machines for predicting/preventing student drop out [145]; feed-forward neural networks, support vector machines, and a probabilistic ensemble simplified fuzzy ARTMAP algorithm to predict dropouts in e-learning courses [156]; 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) [287]; 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 [187]; and the J48 decision tree algorithm and farthest-first clustering algorithm for predicting, understanding, and preventing academic failure (exam failure) among university students [42].

Different types of clustering also used to carry out this task are: Kohonen nets to detect students that cheat in online assessments [43]; outlier detection to uncover atypical student behavior [265]; an outlier detection method using Bayesian predictive distribution to detect learners' irregular learning [263]; 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 [265].

Finally, other DM techniques and models used for this task are, for example, association-rule mining for selecting weak students for remedial classes [163], to send warning messages to students with unusual learning behavior in an AEHS [133], and to construct concept-effect relationships for diagnosing student's learning problems [126]; 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 [169]; stepwise regression to detect misplay and look for sources of error in the prediction of student's test scores [79]; 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 [109].

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 [229].

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 [256]; model-based clustering to automatically discover useful groups from LMS data to obtain profiles of student's behavior [254]; 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 [294]; discriminating features and external profiling features (pass/fail) to support teachers in collaborative student modeling [90]; an improvement in the matrix-based clustering method for grouping learners by characteristics in e-learning [295]; a fuzzy clustering algorithm to find interested groups of learners according to their personality and learning strategy data collected from an online course [259]; a hybrid method of clustering and Bayesian networks to group students according to their skills [105]; 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's skills [188]; a *K*-means clustering algorithm to discover interesting patterns that characterize the work of stronger and weaker students [209]; a conditional subspace clustering algorithm to identify skills that differentiate students [194]; a two-step cluster analysis to classify how students organize personal information spaces (piling, one-folder, small-folders, and big-folder filing) [108]; hierarchical cluster analysis to establish the proportion of students who get an exercise wrong or right [24]; and 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) [304].

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) [252]; classification and regression tree, chi-squared automatic interaction detection, and C4.5 algorithm for the automatic identification of the students' cognitive styles [153]; 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 [151]; a hidden-Markov-model-based classification approach to characterize different types of users through their navigation or content access patterns [85]; decision trees for classifying students according to their accumulated knowledge in e-learning systems [179]; 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 [138]; 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 [290]; decision tree and support vector machine for assessing an activity by more than one lecturer using a pairwise learning model [210]; a classification algorithm for speech act patterns to assess participants' roles and identify discussion threads [141]; 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 [87]. 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) [113]. Collaborative filtering systems can produce personal recommendations by computing the similarity between students' preferences; therefore, this task is directly related to the previous task of recommendations for students (see Section II-F).

Collaborative filtering has been used for context-aware LO recommendation lists [154]; to make a recommendation for a learner about what he/she should learn before taking the next step [300]; 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 [285]; for recommending relevant links to the active learner [147]; to develop an e-learning recommendation service system [157]; and to find relevant content on the Web, personalizing and adapting this content to learners [257].

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 in revealing aspects of their communication [233]. DM and social networks have been used to analyze the structure and content of educative online communities [216]. 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 [190]. 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) [206]. 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 [231].

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 [193]. 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 [262]; to discover concept-effect relationships for diagnosing the learning problems of students [126]; and for conceptual diagnosis of e-learning through automatically constructed concept maps that enable teachers to overcome the learning barrier and misconceptions of learners [152].

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 [149]; to find relationships between text documents and construct document index graphs [107]; and to explore cognitive concept-map differences in instructional outcomes [119].

Finally, a specific concept-map algorithm has been created to automatically organize knowledge points and map them [243]; a method of automatic concept relationship discovery for an adaptive e-course has been developed to help teachers to author overall automation [245]; and a multiexpert 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 [255]. Rough set theory and clustering concept hierarchy have been used to construct e-learning frequently asked questions (FAQ) retrieval infrastructures [56].Multilingual knowledge-discovery technique processing has been combined with AH techniques to automatically create online information systems from linear texts in electronic format, such as textbooks [3]. Argument mining has been proposed to support argument construction for agents and ITSs using different mining techniques [1].

Several DM techniques have been applied to reuse learning resources. Hybrid unsupervised DM techniques have been employed to facilitate LO reuse and retrieval from the Web or from different LO repositories [142]. Valuable information can be found by mining metadata from educational resources (ontology of pedagogical objects), which helps DM to retrieve more precise information for content reuse and exchange [175]. 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 [235]. Profile analysis based on collaborative filtering has been used to search LOs and rank search results according to the predicted level of user interest [197]. Mining educational multimedia presentations has been used to establish explicit relationships among the data related to interactivity (links and actions) and to help to 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 [162]. 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 [118]. Classification, prediction, association-rule analysis, clustering, etc., have been compared to discover new explicit knowledge that 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 [121]. Decision trees and Bayesian models have been proposed to help management institutes to explore the probable effects of changes in recruitments, admissions, and courses [215].

Association-rule mining has been used to provide new, important, and therefore, demand-oriented impulses for the development of new bachelor and master courses [237]. 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 in making decisions on how to improve the quality of the service provided by the university based on students' success and failure rates [239]. 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 [278].

Finally, a regression model has been developed to predict the likelihood a specific undergraduate applicant has of matriculation, if admitted [139]; several clustering algorithms (self-organizing map networks, *K*-means, and *K*th-nearest neighbor) have been used as a decision support in selecting Association of Advance Collegiate Schools of Business (AACSB) peer schools [140].

III. Future Work and Research Lines

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

1) EDM tools have to be designed to be easier for educators or nonexpert users in DM. DM tools are normally designed more for power and flexibility than for simplicity. Most

of the current DM 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 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 DM 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 therefore, 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 DM algorithms to simplify the configuration and execution for nonexpert 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 [93]. 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.

2) Integration with the e-learning system. The DM tool has to be integrated into the e-learning environment as one more traditional authoring tool (course creator, test creator, report tools, etc.). All DM tasks (preprocessing, DM, 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 DM techniques could be easily and directly applied to the e-learning environment using an iterative evaluation process [224].
3) Standardization of data and models. Current tools for mining data pertaining to a specific course/framework may be useful to their developers only. There are no general tools or reusing tools that can be applied to any educational system. Therefore, a standardization of input data and output model are needed, as along with preprocessing, discovering, and postprocessing tasks. Shen *et al.* [243] proposed using Extensible Markup Language (XML) as data specification. Ventura *et al.* [267] used Predictive Modeling Markup Language (PMML) that is the leading standard for statistical and DM models. But, it is also necessary to incorporate domain knowledge and semantics using ontology-specification languages, such as Ontology Web Language (OWL) and Resource Description Framework (RDF); and standard metadata for e-learning, such as SCORM. In this line, currently, there is only one public educational data repository, the PSLC DataShop [143], which provides a lot of educational datasets and also facilitates analysis. However, all this log data are obtained from ITSs; therefore, 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 DM algorithms.
4) Traditional mining algorithms need to be tuned to take into account the educational context. DM 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 DM algorithms. For example, Iksal and Choquet [129] have proposed specific usage tracking language (UTL) to describe the track semantics recorded by an LMS 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. In fact, after first collecting and consulting all the published bibliography in EDM area, we have selected each author's most important studies. Then, we have classified each study 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 upcoming research area related to several well-established areas of research, including e-learning, AH, ITSs, WM, DM, 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 DM algorithms in educational data/environments. Therefore, it could be said that EDM is now approaching its adolescence, i.e., 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 DM 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.

References

[1] S. Abbas and H. Sawamura, "A first step towards argument mining and its use in arguing agents and ITS," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, Zagreb, Croatia, 2008, pp. 149–157.

[2] F. Abel, I. I. Bittencourt, N. Henze, D. Krause, and J. Vassileva, "A rule-based recommender system for online discussion forums," in *Proc. Int. Conf. Adaptive Hypermedia Adaptive Web-Based Syst.*, Hannover, Germany, 2008, pp. 12–21.

[3] E. Alfonseca, P. Rodriguez, and D. Perez, "An approach for automatic generation of adaptive hypermedia in education with multilingual knowledge discovery techniques," *Comput. Educ. J.*, vol. 49, no. 2, pp. 495–513, 2007.

[4] S. Amershi and C. Conati, "Combining unsupervised and supervised classification to build user models for exploratory learning environments," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 18–71, 2009.

[5] A. Anaya and J. Boticario, "A data mining approach to reveal representative collaboration indicators in open collaboration frameworks," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 210–218.

[6] A. Andrejko, M. Barla, M. Bielikova, and M. Tvarozek, "User characteristics acquisition from logs with semantics," in *Proc. Int. Conf. Inf. Syst. Implementation Model.*, Czech Republic, 2007, pp. 103–110.

[7] N. Anozie and B. W. Junker, "Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring

system," in *Proc. AAAI Workshop Educ. Data Mining*, Menlo Park, CA, 2006, pp. 1–6.

[8] A. Arnold, R. Scheines, J. E. Beck, and B. Jerome, "Time and attention: Students, sessions, and tasks," in *Proc. AAAI 2005 Workshop Educ. Data Mining*, Pittsburgh, PA, pp. 62–66.

[9] C. Antunes, "Acquiring background knowledge for intelligent tutoring systems," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 18–27.

[10] I. Arroyo, T. Murray, and B. P. Woolf, "Inferring unobservable learning variables from students' help seeking behavior," in *Proc. Int. Conf. Intell. Tutoring Syst.*, Brazil, 2004, pp. 782–784.

[11] N. Avouris, V. Komis, G. Fiotakis, M. Margaritis, and E. Voyiatzaki, "Why logging of fingertip actions is not enough for analysis of learning activities," in *Proc. AIED Conf. Workshop Usage Anal. Learn. Syst.*, Amsterdam, The Netherlands, 2005, pp. 1–8.

[12] E. Ayers and B. W. Junker, "Do skills combine additively to predict task difficulty in eighth grade mathematics?" in *Proc. AAAI Workshop Educ. Data Mining*, Menlo Park, CA: AAAI Press, 2006, pp. 14–20.

[13] M. Bari and B. Lavoie, "Predicting interactive properties by mining educational multimedia presentations," in *Proc. Int. Conf. Inf. Commun. Technol.*, 2007, pp. 231–234.

[14] E. Ayers, R. Nugent, and N. Dean, "A comparison of student skill knowledge estimates," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 1–10.

[15] R. Baker, A. Corbett, and K. Koedinger, "Detecting student misuse of intelligent tutoring systems," in *Proc. Int. Conf. Intell. Tutoring Syst.*, Alagoas, Brazil, 2004, pp. 531–540.

[16] R. Baker, "Modeling and understanding students' off-task behavior in intelligent tutoring systems," in *Proc. Conf. Hum. Factors Comput. Syst.*, San Jose, CA, 2007, pp. 1059–1068.

[17] R. Baker, J. E. Beck, B. Berendt, A. Kroner, E. Menasalvas, and S. Weibelzahl, "Track on educational data mining," presented at the 11th Int. Conf. User Model. Workshop Data Mining User Model, Corfu, Greece, 2007.

[18] R. Baker, A. T. Corbett, and V. Aleven, "Improving contextual models of guessing and slipping with a truncated training set," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 67–76.

[19] R. Baker, T. Barnes, and J. E. Beck, presented at the 1st Int. Conf. Educ. Data Mining, Montreal, QC, Canada, 2008.

[20] R. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[21] R. Baker, "Data mining for education," in *International Encyclopedia of Education*, B. McGaw, P. Peterson, and E. Baker, Eds., 3rd ed. Oxford, U.K.: Elsevier, 2010.

[22] R. Baker, A. Merceron, and P. I. Pavilk, presented at the 3rd Int. Conf. Educ. Data Mining, Pittsburgh, PA, 2010.

[23] H. Ba-Omar, I. Petrounias, and F. Anwar, "A framework for using web usage mining for personalise e-learning," in *Proc. Int. Conf. Adv. Learn. Technol.*, Niigata, Japan, 2007, pp. 937–938.

[24] D. Barker-Plummer, R. Cox, and R. Dale, "Dimensions of difficulty in translating natural language into fist order logic," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 220–228.

[25] T. Barnes, "The q-matrix method: Mining student response data for knowledge," in *Proc. AAAI Workshop Educ. Data Mining*, Pittsburgh, PA, 2005, pp. 1–8.

[26] T. Barnes and J. Stamper, "Toward automatic hint generation for logic proof tutoring using historical student data," in *Int. Conf. Intell. Tutoring Syst.*, Montreal, QC, Canada, 2008, pp. 373–382.

[27] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, presented at the 2nd Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009.

[28] C. B. Baruque, M. A. Amaral, A. Barcellos, J. C. Da Silva Freitas, and C. J. Longo, "Analysing users' access logs in Moodle to improve e learning," in *Proc. Euro Amer. Conf. Telematics Inf. Syst.*, Faro, Portugal, 2007, pp. 1–4.

[29] C. R. Beal and P. R. Cohen, "Temporal data mining for educational applications," in *Proc. 10th Pacific Rim Int. Conf. Artif. Intell.: Trends Artif. Intell.*, Hanoi, Vietnam, 2008, pp. 66–77.

[30] J. E. Beck, presented at the 5th Int. Conf. Intell. Tutoring Syst. (ITS) Workshop Applying Mach. Learning to ITS Design/Construction, Montreal, QC, Canada, 2000.

[31] J. E. Beck, and B. P. Woolf, "High-level student modeling with machine learning," in *Proc. 5th Int. Conf. Intell. Tutoring Syst.*, Alagoas, Brazil, 2000, pp. 584–593.

[32] J. E. Beck, R. Baker, A. T. Corbett, J. Kay, D. J. Litman, T. Mitrovic, and S. Ritter, presented at the 7th Int. Conf. Workshop Analyzing Student-tutor Interaction Logs Improve Educ. Outcomes, Alagoas, Brazil, 2004.

[33] J. E. Beck, presented at the 20th Nat. Conf. Artif. Intell. (AAAI) Workshop Educ. Data Mining, Pittsburgh, PA, 2005.

[34] J. E. Beck, E. Aimeur, and T. Barnes, presented at the 21st Nat. Conf. Artif. Intell. (AAAI) Workshop Educ. Data Mining, Boston, MA, 2006.

[35] J. E. Beck, M. Pechenizkiy, T. Calders, and S. R. Viola, presented at the 7th IEEE Int. Conf. Adv. Learn. Technol. Workshop Educ. Data Mining, Niigata, Japan, 2007.

[36] K. Becker, C. Ghedini, and E. Terra, "Using kdd to analyze the impact of curriculum revisions in a Brazilian university," in *Proc. 11th Int. Conf. Data Eng.*, Orlando, FL, 2000, pp. 412–419.

[37] A. Bellaachia and E. Vommina, "MINEL: A framework for mining e-learning logs," in *Proc. Fifth IASTED Int. Conf. Web-Based Educ.*, Mexico, 2006, pp. 259–263.

[38] D. Ben-naim, N. Marcus, and M. Bain, "Visualization and analysis of student interaction in an adaptive exploratory learning environment," in *Proc. Int. Workshop Intell. Support Exploratory Environ. Eur. Conf. Technol. Enhanced Learn.*, Maastricht, The Netherlands, 2008, pp. 1–10.

[39] D. Ben-naim, M. Bain, and N. Marcus, "A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 21–30.

[40] L. Burr and D. H. Spennemann, "Pattern of user behavior in university online forums," *Int. J. Instruct. Technol. Distance Learn.*, vol. 1, no. 10, pp. 11–28, 2004.

[41] J. Bravo and A. Ortigosa, "Detecting symptoms of low performance using production rules," presented at the Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009.

[42] V. P. Bresfelean, M. Bresfelean, and N. Ghisoiu, "Determining students' academic failure profile founded on data mining methods," in *Proc. Int. Conf. Inf. Technol. Interfaces*, Croatia, 2008, pp. 317–322.

[43] G. Burlak, J. Muñoz, A. Ochoa, and J. A. Hernández, "Detecting cheats in online student assessments using data mining," in *Proc. Int. Conf. Data Mining*, Las Vegas, NV, 2006, pp. 204–210.

[44] M. Cakir, F. Xhafa, N. Zhou, and G. Stahl, "Thread-based analysis of patterns of collaborative interaction in chat," in *Proc. Int. Conf. AI Educ.*, Amsterdam, The Netherlands, 2005, pp. 121–127.

[45] C. C. Chan, "A framework for assessing usage of web-based e-learning systems," in *Proc. Int. Conf. Innovative Comput., Inf. Control*, Washington, DC, 2007, pp. 147–151.

[46] F. H. Chanchary, I. Haque, and M. S. Khalid, "Web usage mining to evaluate the transfer of learning in a web-based learning environment," in *Proc. Int. Workshop Knowl. Discov. Data Mining*, Washington, DC, 2008, pp. 249–253.

[47] K. M. Chang, J. E. Beck, J. Mostow, and A. Corbett, "A bayes net toolkit for student modeling in intelligent tutoring systems," in *Proc. Int. Conf. Intell. Tutoring Syst.*, Jhongli, Taiwan, 2006, pp. 104–113.

[48] Y. C. Chang, W. Y. Kao, C. P. Chu, and C. H. Chiu, "A learning style classification mechanism for e-learning," *Comput. Educ. J.*, vol. 53, no. 2, pp. 273–285, 2009.

[49] G. Chen, C. Liu, K. Ou, and B. Liu, "Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology," *J. Educ. Comput. Res.*, vol. 23, no. 3, pp. 305–332, 2000.

[50] C. Chen, L. Duh, and C. Liu, "A personalized courseware recommendation system based on fuzzy item response theory," in *Proc. IEEE Int. Conf. E-Technol., E-Commerce E-Service*, Washington, DC, 2004, pp. 305–308.

[51] C. Chen, M. Chen, and Y. Li, "Mining key formative assessment rules based on learner portfiles for web-based learning systems," in *Proc. IEEE Int. Conf. Adv. Learn. Technol.*, Niigata, Japan, 2007, pp. 1–5.

[52] N. S. Chen, Kinshuk, C. W. Wei, and H. J. Chen, "Mining e-learning domain concept map from academic articles," *Comput. Educ. J.*, vol. 50, pp. 1009–1021, 2008.

[53] C. H. Chen, C. M. Hong, and C. C. Chang, "Mining interactive social network for recommending appropriate learning partners in a Web-based cooperative learning environment," in *Proc. IEEE Conf. Cybern. Intell. Syst.*, Chengdu, China, 2008, pp. 642–647.

[54] Y. Chen and C. Weng, "Mining fuzzy association rules from questionnaire data," *Knowl.-Based Syst. J.*, vol. 22, no. 1, pp. 46–56, 2009.

[55] C. Chen and M. Chen, "Mobile formative assessment tool based on data mining techniques for supporting web-based learning," *Comput. Educ. J.*, vol. 52, no. 1, pp. 256–273, 2009.

[56] D. Y. Chiu, Y. C. Pan, and W. C. Chang, "Using rough set theory to construct e-learning faq retrieval infrastructure," in *Proc. IEEE Ubi-Media Comput. Conf.*, Lanzhou, China, 2008, pp. 547–552.

[57] H. C. Chu, G. J. Hwang, J. C. R. Tseng, and G. H. Hwang, "A computerized approach to diagnosing student learning problems in health education," *Asian J. Health Inf. Sci.*, vol. 1, no. 1, pp. 43–60, 2006.

[58] H. C. Chu, G. J. Hwang, P. H. Wu, and J. M. Chen, "A computer-assisted collaborative approach for E-training course design," in *Proc. IEEE Conf. Adv. Learn. Technol.*, Niigata, Japan, 2007, pp. 36–40.

[59] F. Castro, A. Vellido, A. Nebot, and J. Minguillon, "Detecting atypical student behaviour on an e-learning system," in *Proc. Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación*, Granda, Spain, 2005, pp. 153–160.

[60] F. Castro, A. Vellido, A. Nebot, and F. Mugica, "Applying data mining techniques to e-learning problems," in *Evolution of Teaching and Learning Paradigms in Intelligent Environment* (Studies in Computational Intelligence), vol. 62, L. C. Jain, R. Tedman, and D. Tedman, Eds. New York: Springer-Verlag, 2007, pp. 183–221.

[61] A. Cetintas, L. Si, Y. P. Xin, and C. Hord, "Predicting correctness of problem solving from low-level log data in intelligent tutoring systems," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 230–238.

[62] C. Choquet, V. Luengo, and K. Yacef, presented at the Artif. Intell. Educ. Conf. (AIED) Workshop Usage Analysis Learning Syst., Amsterdam, The Netherlands, 2005.

[63] M. Cocea and S. Weibelzahl, "Can log files analysis estimate learners' level of motivation?" in *Proc. Workshop week Lernen—Wissensentdeckung—Adaptivität*, Hildesheim, Germany, 2006, pp. 32–35.

[64] M. Cocea and S. Weibelzahl, "Cross-system validation of engagement prediction from log files," in *Proc. Int. Conf. Conf. Technol. Enhanced Learn.*, Crete, Greece, 2007, pp. 14–25.

[65] R. M. Crespo, A. Pardo, J. P. Pérez, and C. D. Kloos, "An algorithm for peer review matching using student profiles based on fuzzy classification and genetic algorithms," in *Proc. Int. Conf. Innov. Appl. Artif. Intell.*, Bari, Italy, 2005, pp. 685–694.

[66] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, "Predicting students drop out: A case study," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 41–50.

[67] M. Delgado, E. Gibaja, M. C. Pegalajar, and O. Pérez, "Predicting students' marks from. Moodle logs using neural network models," in *Proc. Int. Conf. Current Dev. Technol.-Assist. Educ.*, Sevilla, Spain, 2006, pp. 586–590.

[68] N. Delavari, S. Phon-amnuaisuk, and M. Beikzadeh, "Data mining application in higher learning institutions," *Inf. Educ. J.*, vol. 7, no. 1, pp. 31–54, 2008.

[69] M. C. Desmarais, M. Gagnon, and P. Meshkinfram, "Bayesian student models based on item to item knowledge structures," in *Proc. Conf. Technol. Enhanced Learn.*, Crete, Greece, 2006, pp. 1–10.

[70] G. Dong and J. Pei, *Sequence Data Mining*. New York: Springer-Verlag, 2007.

[71] H. Drachsler, H. G. Hummel, and R. Koper, "Personal recommender systems for learners in lifelong learning networks: The requirements, techniques and model," *Int. J. Learn. Technol.*, vol. 3, no. 4, pp. 404–423, 2008.

[72] N. R. Draper and H. Smith, *Applied Regression Analysis*. New York: Wiley, 1998.

[73] L. P. Dringus and T. Ellis, "Using data mining as a strategy for assessing asynchronous discussion forums," *Comput. Educ. J.*, vol. 45, no. 1, pp. 141–160, 2005.

[74] N. El-kechaï and C. Després, "Proposing the underlying causes that lead to the trainnee's erroneous actions to the trainer," in *Proc. Eur. Conf. Technol. Enhanced Learn.*, Crete, Creece, 2007, pp. 41–55.

[75] P. Espejo, S. Ventura, and F. Herrera, "A survey on the application of genetic programming to classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 2, pp. 121–144, 2010.

[76] T. A. Etchells, A. Nebot, A. Vellido, P. J. G. Lisboa, and F. Mugica, "Learning what is important: Feature selection and rule extraction in a virtual course," in *Proc. Eur. Symp. Artif. Neural Netw.*, Bruseles, Belgium, 2006, pp. 401–406.

[77] R. Farzan and P. Brusilovsky, "Social navigation support in a course recommendation system," in *Proc. 4th Int. Conf. Adaptive Hypermedia Adaptive Web-based Syst.*, Dublin, 2006, pp. 91–100.

[78] L. V. Fausett and W. Elwasif, "Predicting performance from test scores using backpropagation and counterpropagation," in *Proc. IEEE World Congr. Comput. Intell.*, Paris, France, 1994, pp. 3398–3402.

[79] M. Feng, N. Heffernan, and K. Koedinger, "Looking for sources of error in predicting student's knowledge," in *Proc. AAAI Workshop Educ. Data Mining*, 2005, pp. 1–8.

[80] M. Feng and N. T. Heffernan, "Informing teachers live about student learning: Reporting in the assistment system," in *Proc. Conf. Artif. Intell. Educ. Workshop Usage Anal. Learn. Syst.*, Amsterdam, The Netherland, 2005, pp. 1–8.

[81] M. Feng and N. Heffernan, "Informing teachers live about student learning: Reporting in the assistment system," *Technol., Instruction, Cognition, Learn. J.*, vol. 3, pp. 1–8, 2006.

[82] M. Feng, N. T. Heffernan, M. Mani, and C. Heffernan, "Using mixed-effects modeling to compare different grain-sized skill models," in *Proc. Workshop Educ. Data Mining*, Menlo Park, CA, 2006, pp. 57–66.

[83] M. Feng and J. Beck, "Back to the future: A non-automated method of constructing transfer models," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 240–248.

[84] M. Feng, J. E. Beck, and N. T. Heffernan, "Using learning decomposition and bootstrapping with randomization to compare the impact of different educational interventions on learning," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 51–60.

[85] A. W. P. Fok, H. S. Wong, and Y. S. Chen, "Hidden markov model based characterization of content access patterns in an e-learning environment," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Amsterdam, The Netherlands, 2005, pp. 201–204.

[86] D. Freedman, R. Purves, and R. Pisani, *Statistics*, 4th ed. New York: Norton, 2007.

[87] L. Freeman, *The Development of Social Network Analysis*. Vancouver, BC, Canada: Empirical Press, 2006.

[88] J. Freyberger, N. T. Heffernan, and C. Ruiz, "Using association rules to guide a search for best fitting transfer models of student learning," in *Proc. Workshop Analyzing Student-Tutor Interaction Logs Improve Educ. Outcomes*, Alagoas, Brazil, 2004, pp. 1–10.

[89] E. Frias-Martinez, S. Chen, and X. Liu, "Survey of data mining approaches to user modeling for adaptive hypermedia," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 6, pp. 734–749.

[90] E. Gaudioso, M. Montero, L. Talavera, and F. Hernandez-Del-Olmo, "Supporting teachers in collaborative student modeling: A framework and an implementation," *Exp. Syst. Appl.*, vol. 36, pp. 2260–2265, 2009.

[91] P. Garcia, A. Amandi, S. Schiaffino, and M. Campo, "Evaluating bayesian networks' precision for detecting student's learning styles," *Comput. Educ. J.*, vol. 49, pp. 794–808, 2007.

[92] E. Garcia, C. Romero, S. Ventura, and C. Castro, "An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering," *User Model. User-Adapted Interaction: J. Personalization Res.*, vol. 19, pp. 99–132, 2009.

[93] E. Garcia, C. Romero, S. Ventura, and C. Castro, "Collaborative data mining tool for education," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 299–306.

[94] T. D. Gedeon and H. S. Turner, "Explaining student grades predicted by a neural network," in *Int. Conf. Neural Netw.*, Nagoya, Japan, 1993, pp. 609–612.

[95] J. Gibbs and M. Rice, "Evaluating student behavior on instructional Web sites using web server logs," in *Proc. Ninth Sloan-C Int. Conf. Online Learn.*, Orlando, FL, 2003, pp. 1–3.

[96] M. Girones and T. A. Fernandez, "Ariadne, a guiding thread in the learning process's labyrinth," in *Proc. Int. Conf. Current Dev. Technol.-Assist. Educ.*, Sevilla, Spain, 2006, pp. 287–290.

[97] P. Golding and O. Donalson, "Predicting academic performance," in *Proc. Frontiers Educ. Conf.*, San Diego, CA, 2006, pp. 21–26.

[98] Y. Gong, D. Rai, J. E. Beck, and N. T. Heffernan, "Does self-discipline impact students' knowledge and learning?," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 61–70.

[99] H. L. Grob, F. Bensberg, and F. Kaderali, "Controlling open source intermediaries—A web log mining approach," in *Proc. Int. Conf. Inf. Technol. Interfaces*, Zagreb, Croatia, 2004, pp. 233–242.

[100] Q. Guo and M. Zhang, "Implement web learning environment based on data mining," *Knowl.-Based Syst. J.*, vol. 22, pp. 439–442, 2009.

[101] S. Ha, S. Bae, and S. Park, "Web mining for distance education," in *Proc. IEEE Int. Conf. Manage. Innovation Technol.*, Singapore, 2000, pp. 715–719.

[102] P. Haddawy, N. Thi, and T. N. Hien, "A decision support system for evaluating international student applications," in *Proc. Frontiers Educ. Conf.*, Milwaukee, WI, 2007, pp. 1–4.

[103] A. F. Hadwin, J. C. Nesbit, D. Jamieson-Noel, J. Code, and P. H. Winne, "Examining trace data to explore self-regulated learning," *Metacognition Learn. J.*, vol. 2, no. 2/3, pp. 107–124, 2007.

[104] M. Hanna, "Data mining in the e-learning domain," *Campus-Wide Inf. Syst.*, vol. 21, no. 1, pp. 29–34, 2004.

[105] W. Hämäläinen, J. Suhonen, E. Sutinen, and H. Toivonen, "Data mining in personalizing distance education courses," in *Proc. World Conf. Open Learn. Distance Educ.*, Hong Kong, 2004, pp. 1–11.

[106] W. Hämäläinen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems," in *Proc. Int. Conf. Intell. Tutoring Syst.*, Taiwan, 2006, pp. 525–534.

[107] K. Hammouda and M. Kamel, "Data Mining in e-learning," in *E-Learning Networked Environments and Architectures: A Knowledge Processing Perspective* (Advanced Information and Knowledge Processing), S. Pierre, Ed. Springer, 2006, pp. 1–28.

[108] S. Hardof-Jaffe, A. Hershkovitz, H. Abu-Kishk, O. Bergman, and R. Nachmias, "How do students organize personal information spaces?" in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 250–258.

[109] E. Heathcote and S. Prakash, "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 *Proc. Int. Conf. Inf. Commun. Technol. Educ.*, Samos Island, Greece, 2007, pp. 1–6.

[110] E. Heathcote and S. Dawson, "Data mining for evaluation, benchmarking and reflective practice in a LMS," in *Proc. World Conf. E-Learn. Corp., Gov., Healthcare Higher Educ.*, Vancouver, BC, Canada, 2005, pp. 326–333.

[111] C. Heiner, R. Baker, and K. Yacef, presented at the 8th Int. Conf. Intell. Tutoring Syst. (ITS) Workshop Educ. Data Mining, Jhongli, Taiwan, 2006.

[112] C. Heiner, N. Heffernan,, and T. Barnes presented at the 13th Int. Conf. Artif. Intell. Educ. Workshop Educ. Data Mining, Los Angeles, CA, 2007.

[113] J. Herlocker, J. Konstan, L. G. Tervin, and J. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst. J.*, vol. 22, no. 1, pp. 5–53, 2004.

[114] J. M. Heraud, L. France, and A. Mille, "Pixed: An ITS that guides students with the help of learners' interaction log," in *Proc. Int. Conf. Intell. Tutoring Syst.*, Maceio, Brazil, 2004, pp. 57–64.

[115] A. Hershkovitz and R. Nachmias, "Developing a log-based motivation measuring tool," in *Proc. 1st Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 226–233.

[116] A. Hershkovitz and R. Nachmias, "Consistency of students' pace in online learning," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 71–80.

[117] N. T. N. Hien and P. Haddawy, "A decision support system for evaluating international student applications," in *Proc. Frontiers Educ. Conf.*, Milwaukee, WI, 2007, pp. 1–6.

[118] T. Hsia, A. Shie, and L. Chen, "Course planning of extension education to meet market demand by using data mining techniques—An example of Chinkuo technology university in Taiwan," *Expert Syst. Appl. J.*, vol. 34, no. 1, pp. 596–602, 2008.

[119] C. Huang, P. Tsai, C. Hsu, and R. Pan, "Exploring cognitive difference in instructional outcomes using text mining technology," in *Proc. IEEE Conf. Syst., Man, Cybern.*, Taipei, Taiwan, 2006, pp. 2116–2120.

[120] J. Huang, A. Zhu, and Q. Luo, "Personality mining method in web based education system using data minig," in *Proc. IEEE Int. Conf. Grey Syst. Intell. Services*, Nanjing, China, 2007, pp. 155–158.

[121] C. Huang, W. Lin, W. Wang, and W. Wang, "Planning of educational training courses by data mining: Using China Motor Corporation as an example," *Expert Syst. Appl. J.*, vol. 36, no. 3, pp. 7199–7209, 2009.

[122] T. Huang, S. Cheng, and Y. Huang, "A blog article recommendation generating mechanism using an SBACPSO algorithm," *Expert Syst. Appl. J.*, vol. 36, no. 7, pp. 10388–10396, 2009.

[123] R. Hübscher, S. Puntambekar, and A. Nye, "Domain specific interactive data mining," in *Proc. 11th Int. Conf. User Model. Workshop Data Mining User Model.*, Corfu, Greece, 2007, pp. 81–90.

[124] T. Hurley and S. Weibelzahl, "Using MotSaRT to support on-line teachers in student motivation," in *Proc. Eur. Conf. Technol. Enhanced Learn.*, Crete, Creece, 2007, pp. 101–111.

[125] W. Y. Hwang, C. B. Chang, and G. J. Chen, "The relationship of learning traits, motivation and performance-learning response dynamics," *Comput. Educ. J.*, vol. 42, pp. 267–287, 2004.

[126] G. J. Hwang, "A data mining approach to diagnosing student learning problems in science courses," *J. Distance Educ. Technol.*, vol. 3, no. 4, pp. 35–50, 2005.

[127] G. J. Hwang, P. S. Tsai, C. C. Tsai, and J. C. R. Tseng, "A novel approach for assisting teachers in analyzing student web-searching behaviors," *Comput. Educ. J.*, vol. 51, pp. 926–938, 2008.

[128] Z. Ibrahim and D. Rusli, "Predicting students' academic performance: Comparing artificial neural network, decision tree and linear regression," in *Proc. Annu. SAS Malaysia Forum*, Kuala Lumpur, Malaysia, 2007, pp. 1–6.

[129] S. Iksal and C. Choquet, "Usage analysis driven by models in a pedagogical context," in *Proc. 12th Int. Conf. Artif. Intell.* Workshop Usage Analysis Learning Syst., Amsterdam, The Netherlands, 2005, pp. 1–8.

[130] A. Ingram, "Using web server logs in evaluating instructional web sites," *J. Educ. Technol. Syst.*, vol. 28, no. 2, pp. 137–157, 1999.

[131] H. Jin, T. Wu, Z. Liu, and J. Yan, "Application of visual data mining in higher-education evaluation system," in *Proc. Int. Workshop Educ. Technol. Comput. Sci.*, Washington, DC, 2009, pp. 101–104.

[132] J. Jovanovic, D. Gasevic, C. Brooks, V. Devedzic, and M. Hatala, "LOCO-Analyist: A tool for raising teacher's awareness in online learning environments," in *Proc. Eur. Conf. Technol. Enhanced Learn.*, Crete, Greece, 2007, pp. 112–126.

[133] B. S. Jong, T. Y. Chan, and Y. L. Wu, "Learning log explorer in e-learning diagnosis," *IEEE Trans. Educ. J.*, vol. 50, no. 3, pp. 216–228, Aug. 2007.

[134] A. Jonsson, M. Hasmik, J. Johns, H. Mehranian, I. Arroyo, B. Woolf, A. Barto, D. Fisher, and S. Mahadevan, "Evaluating the feasibility of learning student models from data," in *Proc. AAAI Workshop Educ. Data Mining*, Pittsburgh, PA, 2005, pp. 1–6.

[135] A. Juan, T. Daradoumis, J. Faulin, and F. Xhafa, "SAMOS: A model for monitoring students' and groups' activities in collaborative e-learning," *Int. J. Learning Technol.*, vol. 4, no. 1–2, pp. 53–72, 2009.

[136] P. Karampiperis and D. Sampson, "Adaptive learning resources sequencing in educational hypermedia systems," *Educ. Technol. Soc. J.*, vol. 8, no. 4, pp. 128–147, 2005.

[137] J. Kay, N. Maisonneuve, K. Yacef, and O. R. Zaiane, "Mining patterns of events in students' teamwork data," in *Proc. Workshop Educ. Data Mining*, Taiwan, 2006, pp. 1–8.

[138] D. Kelly and B. Tangney, "First aid for you: Getting to know your learning style using machine learning," in *Proc. IEEE Int. Conf. Adv. Learning Technol.*, Washington, DC, 2005, p. 1-3.

[139] S. Khajuria, "A model to predict student matriculation from admissions data," M.S. thesis, Ind. Manufact. Syst. Eng., Ohio Univ., Athens, OH, 2007.

[140] M. Y. Kiang, D. M. Fisher, J. V. Chen, S. A. Fisher, and R. T. Chi, "The application of SOM as a decision support tool to identify AACSB peer schools," *Decision Support Syst. J.*, vol. 47, no. 1, pp. 51–59, 2009.

[141] J. Kim, G. Chern, D. Feng, E. Shaw, and E. Hovy, "Mining and assessing discussions on the web through speech act analysis," in *Proc. AAAI Workshop Web Content Mining Human Lang. Technol.*, Athens, GA, 2006, pp. 1–8.

[142] C. C. Kiu and C. S. Lee, "Learning objects reusability and retrieval through ontological sharing: A hybrid unsupervised data mining approach," in *Proc. IEEE Conf. Adv. Learning Technol.*, Niigata, Japan, 2007, pp. 548–550.

[143] K. Koedinger, K. Cunningham, A. Skogsholm, and B. Leber, "An open repository and analysis tools for fine-grained, longitudinal learner data," in *Proc. 1st Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 157–166.

[144] E. M. Kosba, V. Dimitrova, and R. Boyle, "Using student and group models to support teachers in web-based distance education," in *Proc. Int. Conf. User Model.*, Edinburgh, U.K., 2005, pp. 124–133.

[145] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Preventing student dropout in distance learning systems using machine learning techniques," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, Oxford, U.K., 2003, pp. 3–5.

[146] S. B. Kotsiantis and P. E. Pintelas, "Predicting students' marks in Hellenic Open University," in *Proc. IEEE Int. Conf. Adv. Learning Technol.*, Washington, DC, 2005, pp. 664–668.

[147] M. K. Khribi, M. Jemni, and O. Nasraoui, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," in *Proc. IEEE Int. Conf. Adv. Learning Technol.*, Washington, DC, 2008, pp. 241–245.

[148] A. Kristofic, "Recommender system for adaptive hypermedia applications," in *Proc. Stud. Res. Conf. Informat. Inf. Technol.*, Bratislava, Slovakia, 2005, pp. 229–234.

[149] R. Lau, A. Chung, D. Song, and Q. Huang, "Towards fuzzy domain ontology based concept map generation for e-learning," in *Proc. Int. Conf. Web-Based Learning*, Ediburgh, U.K., 2007, pp. 90–101.

[150] C. S. Lee, "Diagnostic, predictive and compositional modeling with data mining in integrated learning environments," *Comput. Educ. J.*, vol. 49, pp. 562–580, 2007.

[151] M. W. Lee, S. Y. Chen, and X. Liu, "Mining learners' behavior in accessing web-based interface," in *Proc. Int. Conf. Edutainment*, Hong Kong, China, 2007, pp. 226–346.

[152] C. H. Lee, G. Lee, and Y. Leu, "Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning," *Expert Syst. Appl. J.*, vol. 36, pp. 1675–1684, 2009.

[153] M. W. Lee, S. Y. Chen, K. Chrysostomou, and X. Liu, "Mining student's behavior in web-based learning programs," *Expert Syst. Appl. J.*, vol. 36, pp. 3459–3464, 2009.

[154] D. Lemire, Boley, H. S. Mcgrath, and M. Ball, "Collaborative filtering and inference rules for context-aware learning object recommendation," *Int. J. Interactive Technol. Smart Educ.*, vol. 2, no. 3, pp. 1–11, 2005.

[155] O. Licchelli, T. M. Basile, D. N. Mauro, F. Esposito, G. Semeraro, and S. Ferilli, "Machine learning approaches for inducing student models," in *Proc. Int. Conf. Innov. Appl. Artif. Intell.*, Ottawa, Canada, 2004, pp. 935–944.

[156] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ. J.*, vol. 53, no. 3, pp. 950–965, 2009.

[157] X. Li, Q. Luo, and J. Yuan, "Personalized recommendation service system in e-learning using web intelligence," in *Proc. 7th Int. Conf. Comput. Sci.*, Beijing, China, 2007, pp. 531–538.

[158] F. Lin, L. Hsieh, and F. Chuang, "Discovering genres of online discussion threads via text mining," *Comput. Educ. J.*, vol. 52, no. 2, pp. 481–495, 2009.

[159] F. Liu and B. Shih, "Learning activity-based e-learning material recommendation system," in *Proc. Int. Symp. Multimedia*, Taichung, Taiwan, 2007, pp. 343–348.

[160] J. Lu, "A personalized e-learning material recommender system," in *Proc. Int. Conf. Inf. Technol. Appl.*, Harbin, China, 2004, pp. 374–379.

[161] F. Lu, X. Li, Q. Liu, Z. Yang, G. Tan, and T. He, "Research on personalized e-learning system using fuzzy set based clustering algorithm," in *Proc. Int. Conf. Comput. Sci.*, Beijing, China, 2007, pp. 587–590.

[162] J. Luan, "Data mining, knowledge management in higher education, potential applications," in *Proc. Workshop Assoc. Inst. Res. Int. Conf.*, Toronto, ON, Canada, 2002, pp. 1–18.

[163] Y. Ma, B. Liu, C. Wong, P. S. Yu, and S. M. Lee, "Targeting the right students using data mining," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD)*, 2000, pp. 457–464.

[164] L. Machado and K. Becker, "Distance education: A web usage mining case study for the evaluation of learning sites," in *Proc. Int. Conf. Adv. Learning Technol.*, Athens, Greece, 2003, pp. 360–361.

[165] T. Madhyastha and E. Hunt, "Mining diagnostic assessment data for concept similarity," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 72–91, 2009.

[166] P. Markellou, I. Mousourouli, S. Spiros, and A. Tsakalidis, "Using semantic web mining technologies for personalized e-learning experiences," in *Proc. Web-Based Educ.*, Grindelwald, Switzerland, 2005, pp. 461–826.

[167] D. Martinez, "Predicting student outcomes using discriminant function analysis," in *Proc. Meeting Res. Plann. Group*, Lake Arrowhead, CA, 2001, pp. 1–22.

[168] N. Matsuda, W. Cohen, J. Sewall, G. Lacerda, and K. R. Koedinger, "Predicting students performance with SimStudent that learns cognitive skills from observation," in *Proc. Int. Conf. Artif. Intell. Educ.*, Amsterdam, The Netherlands, 2007, pp. 467–476.

[169] M. Mavrikis, "Data-driven prediction of the necessity of help requests in ILEs," in *Proc. Int. Conf. Adaptive Hypermedia*, Hannover, Germany, 2008, pp. 316–319.

[170] R. Mazza and D. Vania, "The design of a course data visualizator: An empirical study," in *Proc. Int. Conf. New Educ. Environ.*, Lucerne, Switzerland, 2003, pp. 215–220.

[171] R. Mazza and C. Milani, "GISMO: A graphical interactive student monitoring tool for course management systems," in *Proc. Int. Conf. Technol. Enhanced Learn.*, Milan, Italy, 2004, pp. 1–8.

[172] R. Mazza, *Introduction to Information Visualization*. New York: Springer-Verlag, 2009.

[173] B. Mcdonald, "Predicting student success," *J. Math. Teaching Learning*, vol. 1, pp. 1–14, 2004.

[174] B. M. Mclaren, K. R. Koedinger, M. Schneider, A. Harrer, and L. Lollen, "Bootstrapping novice data: Semi-automated tutor authoring using student log files," in *Proc. Workshop Analyzing Student-Tutor Interaction Logs Improve Educ. Outcomes*, Alagoas, Brazil, 2004, pp. 1–10.

[175] A. Merceron, C. Oliveira, M. Scholl, and C. Ullrich, "Mining for content re-use and exchange - solutions and problems," in *Proc. Int. Conf. Semantic Web Conf.*, Hiroshima, Japan, 2004, pp. 1–2.

[176] A. Merceron and K. Yacef, "Mining student data captured from a web-based tutoring tool: Initial exploration and results," *J. Interactive Learning Res.*, vol. 15, no. 4, pp. 319–346, 2004.

[177] A. Merceron and K. Yacef, "Educational data mining: A case study," in *Proc. Int. Conf. Artif. Intell. Educ.*, Amsterdam, The Netherlands, 2005, pp. 1–8.

[178] A. Merceron and K. Yacef, "Interestingness measures for association rules in educational data," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 57–66.

[179] C. Mihaescu and D. Burdescu, "Testing attribute selection algorithms for classification performance on real data," in *Proc. Int. IEEE Conf. Intell. Syst.*, Varna, Bulgaria, 2006, pp. 581–586.

[180] B. Minaei-bidgoli, D. A. Kashy, G. Kortmeyer, and W. F. Punch, "Predicting student performance: An application of data mining methods with an educational Web-based system," in *Proc. Int. Conf. Frontiers Educ.*, 2003, pp. 13–18.

[181] B. Minaei-bidgoli, P. Tan, and W. Punch, "Mining interesting contrast rules for a web-based educational system," in *Proc. Int. Conf. Mach. Learning Appl.*, Los Angeles, CA, 2004, pp. 1–8.

[182] P. B. Myszkowski, H. Kwasnicka, and U. Markowska-Kaczmar, "Data mining techniques in e-learning celgrid system," in *Proc. Int. Conf. Comput. Inf. Syst. Ind. Manage. Appl.*, Ostrava, The Czech Republic, 2008, pp. 315–319.

[183] D. Monk, "Using data mining for e-learning decision making," *Electron. J. E-Learning*, vol. 3, no. 1, pp. 41–54, 2005.

[184] E. Mor and J. Minguillón, "E-learning personalization based on itineraries and long-term navigational behavior," in *Proc. 13th World Wide Web Conf.*, New York, 2004, pp. 264–265.

[185] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea, and C. Heiner, "An educational data mining tool to browse tutor-student interactions: Time will tell!," in *Proc. Workshop Educ. Data Mining*, 2005, pp. 15–22.

[186] J. Mostow and J. Beck, "Some useful tactics to modify, map and mine data from intelligent tutors," *J. Nat. Lang. Eng.*, vol. 12, no. 2, pp. 195–208, 2006.

[187] M. Muehlenbrok, "Automatic action analysis in an interactive learning environment," in *Proc. 12th Int. Conf. Artif. Intell. Educ. Workshop Usage Anal. Learn. Syst.*, Amsterdam, The Netherlands, 2005, pp. 73–80.

[188] N. Myller, J. Suhonen, and E. Sutinen, "Using data mining for improving web-based course design," in *Proc. Int. Conf. Comput. Educ.*, Washington, DC, 2002, pp. 959–964.

[189] R. Nagata, K. Takeda, K. Suda, J. Kakegawa, and K. Morihiro, "Edu-mining for book recommendation for pupils," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 91–100.

[190] E. Nankani, S. Simoff, S. Denize, and L. Young, "Supporting strategic decision making in an enterprise university through detecting patterns of academic collaboration," in *Proc. Int. United Inf. Syst. Conf.*, Sydney, Australia, 2009, pp. 496–507.

[191] A. Nebot, F. Castro, A. Vellido, and F. Mugica, "Identification of fuzzy models to predict students perfornance in an e-learning environment," in *Proc. Int. Conf. Web-Based Educ.*, Puerto Vallarta, Mexico, 2006, pp. 74–79.

[192] J. C. Nesbit, Y. Xu, P. H. Winne, and M. Zhou, "Sequential pattern analysis software for educational event data," in *Proc. Int. Conf. Methods Tech. Behav. Res.*, Netherlands, 2008.

[193] J. D. Novak and A. J. Cañas, "The theory underlying concept maps and how to construct and use them," Inst. Human Mach. Cogn., Pensacola. FL, Tech. Rep. IHMC CmapTools 2006-01, 2006.

[194] R. Nugent, E. Ayers, and N. Dean, "Conditional subspace clustering of skill mastery: Identifying skills that separate students," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 101–110.

[195] E. N. Ogor, "Student academic performance monitoring and evaluation using data mining techniques," in *Proc. Electron., Robot. Automotive Mech. Conf.*, Washington, DC, 2007, pp. 354–359.

[196] V. O. Oladokun, A. T. Adebanjo, and O. E. Charles-Owaba, "Predicting student's academic performance using artificial neural network: A case study of an engineering course," *Proc. Pacific J. Sci. Technol.*, vol. 9, no. 1, pp. 72–79, 2008.

[197] T. Orzechowski, S. Ernst, and A. Dziech, "Profiled search methods for e-learning systems," in *Proc. Int. Workshop Learning Object Discov.*

Exchange Eur. Conf. Technol. Enhanced Learn., Crete, Greece, 2007, pp. 1–10.

[198] Y. Ouyang and M. Zhu, "eLORM: Learning object relationship mining based repository," in *Proc. IEEE Int. Conf. Enterprise Comput., E-Commerce, E-Service*, Tokyo, Japan, 2007, pp. 691–698.

[199] C. Pahl and C. Donnellan, "Data mining technology for the evaluation of web-based teaching and learning systems," in *Proc. Congr. E-Learning.*, Montreal, Canada, 2003, pp. 1–7.

[200] Z. Pardos, N. Heffernan, B. Anderson, and C. Heffernan, "The effect of model granularity on student performance prediction using bayesian networks," in *Proc. Int. Conf. User Model.*, Corfu, Greece, 2007, pp. 435–439.

[201] Z. Pardos, J. E. Beck, C. Ruiz, and N. Heffernan, "The composition effect: Conjunctive or compensatory? An analysis of multi-skill math questions in ITS," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 147–156.

[202] Z. Pardos, J. E. Beck, and N. Heffernan, "Determining the significance of item order in randomized problem sets," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 111–120.

[203] P. Pavlik, H. Cen, and K. Koedinger, "Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 121–130.

[204] M. Pechenizkiy, T. Calders, E. Vasilyeva, and P. De bra, "Mining the student assessment data: Lessons drawn from a small scale case study," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 187–191.

[205] M. Pechenizkiy, N. Trcka, E. Vasilyeva, W. Aalst, and P. De bra, "Process mining online assessment data," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 279–288.

[206] D. Prata, R. Baker, E. Costa, C. Rose, and Y. Cui, "Detecting and understanding the impact of cognitive and interpersonal conflict in computer supported collaborative learning environments," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 131–140.

[207] D. Pritchard and R. Warnakulasooriya, "Data from a web-based homework tutor can predict student's final exam score," in *Proc. World Conf. Educ. Multimedia, Hypermedia Telecommun.*, Chesapeake, VA, 2005, pp. 2523–2529.

[208] Y. Psaromiligkos, M. Orfanidou, C. Kytagias, and E. Zafiri, "Mining log data for the analysis of learners' behaviour in web-based learning management systems," *Oper. Res. J.*, vol. 11, pp. 1–14, 2009.

[209] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaïane, "Clustering and sequential pattern mining of online collaborative learning data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 6, pp. 759–772, Jun. 2009.

[210] J. R. Quevedo and E. Motañes, "Obtaining rubric weitghts for assessments by more than one lecturer using a pairwise learning model," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 289–298.

[211] S. N. R. Raghavan, "Data mining in e-commerce: A survey," *Sadhana J.*, vol. 30, no. 2/3, pp. 275–289, 2005.

[212] M. Rahkila and M. Karjalainen, "Evaluation of learning in computer based education using log systems," in *Proc. ASEE/IEEE Front. Educ. Conf.*, San Juan, Puerto Rico, 1999, pp. 16–21.

[213] D. Rai, Y. Gong, and J. E. Beck, "Using Dirichlet priors to improve model parameter plausibility," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 141–150.

[214] A. A. Ramli, "Web usage mining using a priori algorithm: UUM learning care portal case," in *Proc. Int. Conf. Knowl. Manage.*, 2005, Malaysia, pp. 1–19.

[215] J. Ranjan and S. Khalil, "Conceptual framework of data mining process in management education in India: An institutional perspective," *Inf. Technol. J.*, vol. 7, no. 1, pp. 16–23, 2008.

[216] R. Rallo, M. Gisbert, and J. Salinas, "Using data mining and social networks to analyze the structure and content of educative online communities," in *Proc. Int. Conf. Multimedia ICTs Educ.*, Caceres, Spain, 2005, pp. 1–10.

[217] S. Ritter, T. Harris, T. Nixon, D. Dickison, R. Murray, and B. Towle, "Reducing the knowledge tracing space," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 151–160.

[218] V. Robinet, G. Bisson, M. Gordon, and B. Lemaire, "Searching for student intermediate mental steps," in *Proc. Int. Conf. User Model. Workshop Data Mining User Model.*, Corfu, Greece, 2007, pp. 101–105.

[219] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware author," *User Model. User-Adapted Interaction: J. Personalization Res.*, vol. 14, no. 5, p. 425-464, 2004.

[220] C. Romero and S. Ventura, *Data Mining in E-Learning*. Ashurst, Southampton: Wit Press, 2006.

[221] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 1, no. 33, pp. 135–146, 2007.

[222] C. Romero, M. Pechenizkiy, T. Calders, and S. R. Viola, in *Proc. Int. Workshop Applying Data Mining E-Learning (ADML): 2nd Eur. Conf. Technol. Enhanced Learn. (EC-TEL'07)*, Crete, Creece, 2007.

[223] C. Romero, S. Ventura, and E. Salcines, "Data mining in course management systems: Moodle case study and tutorial," *Comput. Educ.*, vol. 51, no. 1, pp. 368–384, 2008.

[224] C. Romero, S. ventura, C. hervás, and P. Gonzales, "Data mining algorithms to classify students," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, Canada, 2008, pp. 8–17.

[225] C. Romero, S. Gutierrez, M. Freire, and S. Ventura, "Mining and visualizing visited trails in web-based educational systems," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, Canada, 2008, pp. 182–185.

[226] C. Romero, P. Gonzalez, S. Ventura, M. J. del Jesus, and F. Herrera, "Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data," *Expert Syst. Appl. J.*, vol. 36, pp. 1632–1644, 2009.

[227] C. Romero, S. Ventura, A. Zafra, and P. de bra, "Applying web usage mining for personalizing hyperlinks in web-based adaptive educ. systems," *Comput. Educ.*, vol. 53, no. 3, pp. 828–840, 2009.

[228] C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, *Handbook of Educational Data Mining*. New York: Taylor & Francis, 2010.

[229] H. C. Romesburg, *Cluster Analysis for Researchers*. Melbourne, FL: Krieger, 2004.

[230] F. Rosta and P. Brusilovsky, "Social navigation support in a course recommendation system," in *Proc. Int. Conf. Adaptive Hypermedia Adaptive Web-Based Syst.*, Dublin, Ireland, 2006, pp. 1–10.

[231] C. Reffay and T. Chanier, "How social network analysis can help to measure cohesion in collaborative distance-learning," in *Proc. Int. Conf. Comput. Supported Collaborative Learning*, Bergen, Norvège, 2003, pp. 1–6.

[232] S. Retalis, A. Papasalouros, Y. Psaromilogkos, S. Siscos, and T. Kargidis, "Towards networked learning analytics—A concept and a tool," in *Proc. 5th Int. Conf. Netw. Learning*, 2006, pp. 1–8.

[233] P. Reyes and P. Tchounikine, "Mining learning groups' activities in forum-type tools," in *Proc. Conf. Comput. Support Collaborative Learning: Learning*, Taipei, Taiwan, 2005, pp. 509–513.

[234] V. Rus, M. Lintean, and R. Azevedo, "Automatic detection of student mental models during prior knowledge activation in MetaTutor," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 161–170.

[235] P. S. Saini, D. Sona, S. Veeramachaneni, and M. Ronchetti, "Making e-learning better through machine learning," in *Proc. Int. Conf. Methods Technol. Learning*, Barcelona, Spain, 2005, pp. 1–6.

[236] A. P. Sanjeev and J. M. Zytkow, "Discovering enrollment knowledge in university databases," in *Proc. Int. Conf. Knowl. Discov. Data Mining*, Montreal, Canada, 1995, pp. 246–251.

[237] K. Schönbrunn and A. Hilbert, "Data mining in higher education," in *Proc. Annu. Conf. Gesellschaft für Klassifikation e.V., Freie Universität Berlin*, 2007, pp. 489–496.

[238] J. Schoonenboom, K. Heller, M. Neenoy, and M. Levene, *Trails in Education: Technologies That Support Navigational Learning*. Rotterdam, The Netherlands: Sense Publisher, 2007.

[239] N. Selmoune and Z. Alimazighi, "A decisional tool for quality improvement in higher education," in *Proc. Int. Conf. Inf. Commun. Technol.*, Damascus, Syria, 2008, pp. 1–6.

[240] D. Shangping and Z. Ping, "A data mining algorithm in distance learning," in *Proc. Int. Conf. Comput. Supported Cooperative Work in Design*, Xian, China, 2008, pp. 1014–1017.

[241] J. Sheard, J. Ceddia, J. Hurst, and J. Tuovinen, "Inferring student learning behaviour from website interactions: A usage analysis," *J. Educ. Inf. Technol.*, vol. 8, no. 3, pp. 245–266, 2003.

[242] R. Shen, F. Yang, and P. Han, "Data analysis center based on e-learning platform," in *Proc. Workshop Internet Challenge: Technol. Appl.*, Berlin, Germany, 2002, pp. 19–28.

[243] R. Shen, P. Han, F. Yang, Q. Yang, and J. Huang, "Data mining and case-based reasoning for distance learning," *J. Distance Educ. Technol.*, vol. 1, no. 3, pp. 46–58, 2003.

[244] L. Shen and R. Shen, "Learning content recommendation service based-on simple sequencing specification," in *Proc. Int. Conf. Web-based Learning*, Beijing, China, 2004, p. 363-370.

[245] M. Simko and M. Bielikova, "Automatic concept relationships discovery for an adaptive e-course," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 171–178.

[246] M. K. Singley and R. B. Lam, “The classroom sentinel: Supporting data-driven decision-making in the classroom,” in *Proc. 13th World Wide Web Conf.*, Chiba, Japan, 2005, pp. 315–322.
[247] D. Song, H. Lin, and Z. Yang, “Opinion mining in e-learning system,” in *Proc. Int. Conf. Netw. Parallel Comput. Workshops*, Washington, DC, 2007, pp. 788–792.
[248] J. Spacco, T. Winters, and T. Payne, “Inferring use cases from unit testing,” in *Proc. Workshop Educ. Data Mining*, New York, 2006, pp. 1–7.
[249] J. Stamper and T. Barnes, “Unsupervised MDP value selection for automating ITS capabilities,” in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 180–188.
[250] R. Stevens, A. Giordani, M. Cooper, A. Soller, L. Gerosa, and C. Cox, “Developing a framework for integrating prior problem solving and knowledge sharing histories of a group to predict future group performance,” in *Proc. Int. Conf. Collaborative Comput.: Netw., Appl. Worksharing*, Boston, MA, 2005, pp. 1–9.
[251] Z. Su, W. Song, M. Lin, and J. Li, “Web text clustering for personalized e-learning based on maximal frequent item sets,” in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Washington, DC, 2008, pp. 452–455.
[252] J. F. Superby, J. P. Vandamme, and N. Meskens, “Determination of factors influencing the achievement of the first-year university students using data mining methods,” in *Proc. Int. Conf. Intell. Tutoring Syst. Workshop Educ. Data Mining*, Taiwan, 2006, pp. 1–8.
[253] D. W. Tai, H. J. Wu, and P. H. Li, “Effective e-learning recommendation system based on self-organizing maps and association mining,” *Electron. Library J.*, vol. 26, no. 3, pp. 329–344, 2008.
[254] L. Talavera and E. Gaudioso, “Mining student data to characterize similar behavior groups in unstructured collaboration spaces,” in *Proc. Workshop Artif. Intell. CSCL*, Valencia, Spain, 2004, pp. 17–23.
[255] C. Tang, R. W. H. Lau, Q. Li, H. Yin, T. Li, and D. Kilis, “Personalized courseware construction based on web data mining,” in *Proc. 1st Int. Conf. Web Inf. Syst. Eng.*, Hong Kong, China, 2000, pp. 204–211.
[256] T. Y. Tang and G. Mccalla, “Student modeling for a web-based learning environment: A data mining approach,” in *Proc. Conf. Artif. Intell.*, Edmonton, Canada, 2002, pp. 967–968.
[257] T. Tang and G. Mccalla, “Smart recommendation for an evolving e-learning system,” *Int. J. E-Learning*, vol. 4, no. 1, pp. 105–129, 2005.
[258] E. H. Thomas and N. Galambos, “What satisfies students? Mining student-opinion data with regression and decision tree analysis,” *Res. Higher Educ. J.*, vol. 45, no. 3, pp. 251–269, 2004.
[259] F. Tian, S. Wang, C. Zheng, and Q. Zheng, “Research on e-learning personality group based on fuzzy clustering analysis,” in *Proc. Int. Conf. Comput. Supported Cooperative Work Design*, Xian, China, 2008, pp. 1035–1040.
[260] C. J. Tsai, S. S. Tseng, and C. Y. Lin, “A two-phase fuzzy mining and learning algorithm for adaptive learning environment,” in *Proc. Int. Conf. Comput. Sci.*, San Francisco, 2001, p. 429-438.
[261] L. Tsantis and J. Castellani, “Enhancing learning environments through solution-based knowledge discovery tools,” *J. Spec. Educ. Technol.*, vol. 16, no. 4, pp. 39–52, 2001.
[262] S. S. Tseng, P. C. Sue, J. M. Su, J. F. Weng, and W. N. Tsai, “A new approach for constructing the concept map,” *Comput. Educ. J.*, vol. 49, pp. 691–707, 2007.
[263] M. Ueno and K. Nagaoka, “Learning log database and data mining system for e-learning—on line statistical outlier detection of irregular learning processes,” in *Proc. Int. Conf. Adv. Learning Technol.*, Tatarstan, Russia, 2002, pp. 436–438.
[264] M. Ueno, “Data mining and text mining technologies for collaborative learning in an ILMS "Samurai",” in *Proc. IEEE Int. Conf. Adv. Learning Technol.*, Washington, DC, 2004, pp. 1052–1053.
[265] M. N. Vee, B. Meyer, and K. L. Mannock, “Understanding novice errors and error paths in Object-oriented programming through log analysis,” in *Proc. Workshop Educ. Data Mining*, Taiwan, 2006, pp. 13–20.
[266] A. Vellido, F. Castro, T. A. Etchells, A. Nebot, and F. Mugica, “Data mining of virtual campus data,” in *Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence (SCI)* (Advanced Information and Knowledge Processing), vol. 62. New York: Springer-Verlag, 2007, pp. 223–254.
[267] S. Ventura, C. Romero, and C. Hervas, “Analyzing rule evaluation measures with educational datasets: A framework to help the teacher,” in *Proc. Int. Conf. Educ. Data Mining*, Montreal, Canada, 2008, pp. 177–181.
[268] C. Vialardi, J. Bravo, and A. Ortigosa, “Improving AEH courses through log analysis,” *J. Universal Comput. Sci.*, vol. 14, no. 17, pp. 2777–1798, 2008.
[269] C. Vialardi, J. Bravo, L. Shafti, and A. Ortigosa, “Recommendation in higher education using data mining techniques,” in *Proc. Int. Conf. Educ. Conf.*, Cordoba, Spain, 2009, pp. 190–198.
[270] S. R. Viola, S. Graf, Kinshuk, and T. Leo, “Analysis of Felder-Silverman index of learning styles by a data-driven statistical approach,” in *Proc. 8th IEEE Int. Symp. Multimedia*, Washington, DC, 2006, pp. 959–964.
[271] M. Vranic, D. Pintar, and Z. Skocir, “The use of data mining in education environment,” in *Proc. Int. Conf. Telecommun.*, Zagred, Croatia, 2007, pp. 243–250.
[272] F. H. Wang, “On using data-mining technology for browsing log file analysis in asynchronous learning environment,” in *Proc. World Conf. Educ. Multimedia, Hypermedia Telecommun.*, Chesapeake, VA, 2002, pp. 2005–2006.
[273] F. H. Wang, “A fuzzy neural network for item sequencing in personalized cognitive scaffolding with adaptive formative assessment,” *Expert Syst. Appl. J.*, vol. 27, pp. 11–25, 2004.
[274] F. H. Wang, “Content recommendation based on education-contextualized browsing events for web-based personalized learning,” *Educ. Technol. Soc. J.*, vol. 11, no. 4, pp. 94–112, 2008.
[275] A. Y. Wang and M. H. Newlin, “Predictors of we-based performance: The role olf self-effecacy and reasons for taking an on-line class,” *Comput. Human Behav. J.*, vol. 18, pp. 151–163, 2002.
[276] F. H. Wang and H. M. Shao, “Effective personalized recommendation based on time-framed navigation clustering and association mining,” *Expert Syst. Appl. J.*, vol. 27, pp. 365–377, 2004.
[277] W. Wang, J. Weng, J. Su, and S. Tseng, “Learning portfolio analysis and mining in scorm compliant environment,” in *Proc. ASEE/IEEE Front. Educ. Conf.*, Savannah, GA, 2004, pp. 17–24.
[278] Y. Wang, Y. Cheng, T. Chang, and S. M. Jen, “On the application of data mining technique and genetic algorithm to an automatic course scheduling system,” in *Proc. IEEE Conf. Cybern. Intell. Syst.*, Chengdu, China, 2008, pp. 400–405.
[279] Y. Wang, M. Tseng, and H. Liao, “Data mining for adaptive learning sequence in English language instruction,” *Expert Syst. Appl. J.*, vol. 36, pp. 7681–7686, 2009.
[280] T. Want and A. Mitrovic, “Using neural networks to predict student’s performance,” in *Proc. Int. Conf. Comput. Educ.*, Washington, DC, 2002, pp. 1–5.
[281] T. Winters, C. R. Shelton, T. Payne, and G. Mei, “Topic extraction from item-level grades,” in *Proc. AAAI Workshop Educ. Data Mining*, Pittsburgh, PA, 2005, pp. 7–14.
[282] A. Wu and C. Leung, “Evaluating learning behavior of web-based training (WBT) using web log,” in *Proc. Int. Conf. Comput. Educ.*, New Zealand, 2002, pp. 736–737.
[283] K. Yamanishi and H. Li, “Mining from open answers in questionnaire data,” in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, San Francisco, CA, 2001, pp. 443–449.
[284] T. D. Yang, T. Lin, and K. Wu, “An agent-based recommender system for lesson plan sequencing,” in *Proc. Int. Conf. Adv. Learning Technol.*, Kazan, Russia, 2002, pp. 14–20.
[285] F. Yang, P. Han, R. Shen, and Z. Hu, “A novel resource recommendation system based on connecting to similar e-learners,” in *Proc. Int. Conf. Web-Based Learning*, Hong Kong, China, 2005, pp. 122–130.
[286] J. Yoo, S. Yoo, C. Lance, and J. Hankins, “Student progress monitoring tool using treeview,” in *Proc. Tech. Symp. Comput. Sci. Educ., ACM-SIGCESE*, 2006, pp. 373–377.
[287] M. V. Yudelson, O. Medvedeva, E. Legowski, M. Castine, D. Jukic, and D. Rebecca, “Mining student learning data to develop high level pedagogic strategy in a medical ITS,” in *Proc. AAAI Workshop Educ. Data Mining*, Boston, MA, 2006, pp. 1–8.
[288] C. H. Yu, A. Jannasch-Pennell, S. Digangi, and B. Wasson, “Using online interactive statistics for evaluating web-based instruction,” *J. Educ. Media Int.*, vol. 35, pp. 157–161, 1999.
[289] P. Yu, C. Own, and L. Lin, “On learning behavior analysis of web based interactive environment,” in *Proc. Int. Conf. Comput. Electr. Eng.*, Oslo/Bergen, Norway, 2001, pp. 1–9.
[290] C. H. Yu, S. Digangi, A. K. Jannasch-pennell, and C. Kaprolet, “Profiling students who take online courses using data mining methods,” *Online J. Distance Learning Administ.*, vol. 11, no. 2, pp. 1–14, 2008.
[291] A. Zafra and S. Ventura, “Predicting student grades in learning management systems with multiple instance programming,” in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 307–314.
[292] O. Zaïane and J. Luo, “Web usage mining for a better web-based learning environment,” in *Proc. Conf. Adv. Technol. Educ.*, Banff, AB, Canada, 2001, pp. 60–64.

[293] O. Zaïane, "Building a recommender agent for e-learning systems," in *Proc. Int. Conf. Educ.*, Auckland, New Zealand, 2002, pp. 55–59.

[294] D. Zakrzewska, "Cluster analysis for user's modeling in intelligent e-learning systems," in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.*, Poland, 2008, pp. 209–214.

[295] K. Zhang, L. Cui, H. Wang, and Q. Sui, "An improvement of matrix-based clustering method for grouping learners in e-learning," in *Proc. Int. Conf. Comput. Supported Cooperative Work Design*, Melbourne, Australia, 2007, pp. 1010–1015.

[296] C. Zhang and S. Zhang, *Association Rule Mining: Models and Algorithms* (Lecture Notes in Artificial Intelligence). New York: Springer-Verlag, 2002.

[297] X. Zhang, J. Mostow, N. Duke, C. Trotochaud, J. Valeri, and A. Corbett, "Mining free-form spoken responses to tutor prompts," in *Proc. Int. Conf. Educ. Data Mining*, Montreal, QC, Canada, 2008, pp. 234–241.

[298] L. Zhang, X. Liu, and X. Liu, "Personalized instructing recommendation system based on web mining," in *Proc. Int. Conf. Young Comput. Sci.*, Hunan, China, 2008, pp. 2517–2521.

[299] S. Zheng, S. Xiong, Y. Huang, and S. Wu, "Using methods of association rules mining optimization in web-based mobile learning system," in *Proc. Int. Symp. Electron. Commerce Security*, Guangzhou, China, 2008, pp. 967–970.

[300] F. Zhu, H. Ip, A. Fok, and J. Cao, "PeRES: A personalized recommendation education system based on multi-agents & SCORM," in *Proc. Int. Conf. Web-Based Learning*, Ediburgh, U.K., 2007, pp. 31–42.

[301] C. Zinn and O. Scheuer, "Getting to know your students in distance-learning contexts," in *Proc. 1st Eur. Conf. Tehcnol. Enhanced Learn.*, 2006, pp. 437–451.

[302] L. Zoubek and M. Burda, "Visualization of differences in data measuring mathematical skills," in *Proc. Int. Conf. Educ. Data Mining*, Cordoba, Spain, 2009, pp. 315–324.

[303] M. E. Zorrilla, E. Menasalvas, D. Marin, E. Mora, and J. Segovia, "Web usage mining project for improving web-based learning sites," in *Proc. Int. Conf. Comput. Aided Syst. Theory*, Las Palmas de Gran Canaria, Spain, 2005, pp. 205–210.

[304] Z. Zukhri and K. Omar, "Solving new student allocation problem with genetic algorithms: A hard problem for partition based approach," *J. Zhejiang Univ.*, vol. 2, no. 1, pp. 6–15, 2007.

Cristóbal Romero (M'09) received the B.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 1996 and 2003, respectively.

He is currently an Associate Professor in the Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain. He has authored or coauthored more than 30 papers about educational data mining in international journals and conferences. His research interests include applying data mining in e-learning systems.

Dr. Romero is a member of the IEEE Computer Society, the International Educational Data Mining (EDM) Working Group, and the steering committee of the EDM Conferences.

Sebastián Ventura (SM'07) received the B.Sc. and Ph.D. degrees in sciences from the University of Córdoba, Córdoba, Spain, in 1989 and 1996, respectively.

He is currently an Associate Professor in the Department of Computer Science and Numerical Analysis, University of Córdoba, where he is the Head of the Knowledge Discovery and Intelligent Systems Research Laboratory. He has authored or coauthored 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) that were supported by the Spanish and Andalusian governments and the European Union. His research interests include soft computing, machine learning, data mining, and its applications.

Dr. Ventura is a Senior Member of the IEEE Computer Society, the IEEE Computational Intelligence Society, the IEEE Systems, Man, and Cybernetics Society, and the Association of Computing Machinery.