Chapter 1
Which Contribution Does EDM Provide to Computer-Based Learning Environments?

Nabila Bousbia and Idriss Belamri

Abstract Educational Data Mining is a new growing research area that can be defined as the application of data mining techniques on raw data from educational systems in order to respond to the educational questions and problems, and also to discover the information hidden after this data. Over the last few years, the popularity of this field enhanced a large number of research studies that is difficult to surround and to identify the contribution of data mining techniques in educational systems. In fact, exploit and understand the raw data collected from educational systems can be "a gold mine" to help the designers and the users of these systems improving their performance and extracting useful information on the behaviors of students in the learning process. The use of data mining techniques in e-learning systems could be very interesting to resolve learning problems. Researchers' ambition is to respond to questions like: What can predict learners' success? Which scenario sequence is more efficient for a specific student? What are the student actions that indicate the learning progress? What are the characteristics of a learning environment allowing a better learning? etc. The current feedback allows detecting the usefulness of applying EDM on visualizing and describing the learning raw data. The predictions take also an interest, particularly the prediction of performance and learners' behaviors. The aim of this chapter is to establish a bibliographic review of the various studies made in the field of educational data mining (EDM) to identify the different aspects studied: the analyzed data, the objectives of these studies, the used techniques and the contribution of the application of these techniques in the field of computer based learning. The goal is not only to list the existing work but also to facilitate the use and the understanding of data mining techniques to help the educational field specialists to give their

N. Bousbia (✉) · I. Belamri
Laboratoire Méthode de Conception de Systèmes (LMCS), Ecole nationale Supérieure d'Informatique (ESI), BP 68M, 16309 Oued-Smar, Algiers, Algeria
e-mail: n_bousbia@esi.dz

I. Belamri
e-mail: i_belamri@esi.dz

feedback and to identify promoter research areas in this field to be exploited in the future.

Keywords EDM · Learner's behavior · Prediction of student's performance · Computer based learning environments (CBLE)

Abbreviations

CBLE	Computer based learning environment
DM	Data mining
EDM	Educational data mining
ITS	Intelligent tutoring system
KDD	Knowledge discovery in databases
KT	Knowledge tracing
LA	Learning analytics
LAK	Learning analytics and knowledge
LMS	Learning management system
NMF	Non-negative matrix factorization
SNA	Social network analysis

1.1 Introduction

EDM is an emerging discipline, with a suite of computational and psychological methods and research approaches for understanding how students learn, and the settings which they learn in [1].

Data of interest is not restricted to interactions of individual students with an educational system (e.g., navigation behavior, input to quizzes and interactive exercises) but might also include data from collaborating students (e.g., text chat), administrative data (e.g., school, school district, teacher), demographic data (e.g., gender, age, school grades), and data on student affect (e.g., motivation, emotional states) [2].

EDM can be applied to assess students' learning performance, to improve the learning process and guide students' learning, to provide feedback and adapt learning recommendations based on students' learning behaviors, to evaluate learning materials and courseware, to detect abnormal learning behaviors and problems, and to achieve a deeper understanding of educational phenomena [3].

For example, Ayesha et al. [4] described the use of k-means clustering algorithm to predict student's learning activities. Pal [5] used machine learning algorithm to find students which are likely to drop out their first year of engineering. Parack et al. [6] used multiple data mining algorithms for student profiling and grouping based on their academic records such as exam scores, term work grades, attendance and practical exams.

As the number of EDM studies found in the literature is growing considerably over the last few years, we aim in this chapter to establish a bibliographic review of these studies. Our goal is to discuss the data mining methods and tools used in computer based learning environments to analyze learners' behaviors and performance in order to facilitate the use and the understanding of data mining techniques to help the educational field specialists to give their feedback and to identify promoter research areas in this field to be exploited in the future.

Therefore, the remaining of the chapter is organized as follows: Sect. 1.2 is devoted to give a detailed view of the EDM field: definition, related areas, goals, methods, the analyzed data, process and the used tools. Section 1.3 presents some examples dealing with the two principal EDM applications: analyzing learners' behaviors and predicting learners' performance. We compare and discuss these examples according to their goals, the analyzed data and the used methods. We end the chapter with a conclusion in Sect. 1.4.

1.2 Educational Data Mining

1.2.1 Definition

Different definitions have been provided for the term 'Educational Data Mining' or EDM. Educational data mining is defined by the journal of educational data mining[1] and Baker [1] as "*an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in*".

This definition does not mention data mining; open to exploring and developing other analytical methods that can be applied to educationally related data [7].

However, in [8] the authors precise that: "*EDM is both a learning science, as well as a rich application area for data mining, due to the growing availability of educational data. It enables data-driven decision making for improving the current educational practice and learning material*".

In the same way, Romero and Ventura [9, 10] define EDM as "*the application of data mining (DM) techniques to specific type of dataset that come from educational environments to address important educational questions*".

Although different in some details, these definitions share an emphasis on discovering knowledge based on educational data to improve educational systems. Note also that the definition of EDM is often confused with 'learning analytics' defined on the LAK (Learning Analytics and Knowledge) website as "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs*" [11].

---

[1] http://www.educationaldatamining.org/JEDM/ visited on August 6, 2013.

Fig. 1.1 Areas in relation with EDM [10]
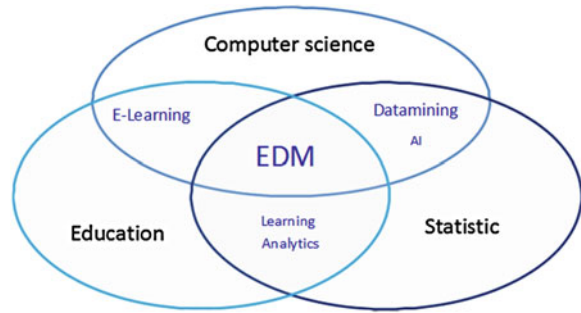
Although there is no hard and fast distinction between these two fields, they have had somewhat different research histories and are developing as distinct research areas [12]. The objective of this chapter is not to draw up a comparative study between these two concepts (comparisons and details can be found in [10–12]).

However, we think worth mentioning that this field is the most related to the EDM field, as they share many goals and it is often difficult to differentiate if an application fits into one or the other of the two areas. The next subsection presents the related fields to EDM.

1.2.2 Areas in Relation to EDM

EDM can be drawn as the combination of three main areas (Fig. 1.1): computer science, education, and statistics. The intersection of those three areas also forms other subareas closely related to EDM such as learning analytics (LA), CBLE, DM and machine learning [10].

As an interdisciplinary area, EDM uses methods and applies techniques from statistics, machine learning, data mining, information retrieval, recommender systems, psycho-pedagogy, cognitive psychology, psychometrics, etc. The choice of which method or technique should be used depends on the addressed educational issue.

1.2.3 Objectives of the EDM

In the last several years, EDM has been applied to address a wide number of goals that are all parts of the general objective of improving learning [10]. Several studies [1, 8, 10, 12, 13] dress a list of these objectives.

Romero and Ventura [10] proposed to classify EDM objectives depending on the viewpoint of the final user (learner, educator, administrator, and researcher) and the problem to resolve:

- Learners. To support a learner's reflections on the situation, to provide adaptive feedback or recommendations to learners, to respond to student's needs, to improve learning performance, etc.
- Educators. To understand their students' learning processes and reflect on their own teaching methods, to improve teaching performance, to understand social, cognitive and behavioral aspects, etc.
- Researchers. To develop and compare data mining techniques to be able to recommend the most useful one for each specific educational task or problem, to evaluate learning effectiveness when using different settings and methods, etc.
- Administrators. To evaluate the best way to organize institutional resources (human and material) and their educational offer.

This view point clearly shows the benefit of EDM applications to the end user, but it is difficult to classify all EDM application goals according to these four actors, especially when an objective is related to more than one actor. That is why, based on the work of [1, 12–14] that focused on the related research goal of EDM applications, we distinguish between the following EDM general goals:

- Student modeling. User modeling in the educational domain incorporates such detailed information as students' characteristics or states such as knowledge, skills, motivation, satisfaction, meta-cognition, attitudes, experiences and learning progress, or certain types of problems that negatively impact their learning outcomes (making too many errors, misusing or under-using help, gaming the system, inefficiently exploring learning resources, etc.), affect, learning styles, and preferences. The common objective here is to create or improve a student model from usage information.
- Predicting students' performance and learning outcomes. The objective is to predict a student's final grades or other types of learning outcomes (such as retention in a degree program or future ability to learn), based on data from course activities. Examples of predicting student's performance can be found in Sect. 1.3.
- Generating recommendation. The objective is to recommend to students which content (or tasks or links) is the most appropriate for them at the current time [15].
- Analyzing learner's behavior. This takes on several forms: Applying educational data mining to answer questions in any of the three areas previously discussed (student models, Prediction, Generating recommendation). It is also used to group student according to their profile, and for adaptation and personalization purposes.
- Communicating to stakeholders. The objective is to help course administrators and educators in analyzing students' activities and usage information in courses. Macfayden and Dawson in [16] conducted a study that confirms that pedagogically meaningful information that is extracted from e-learning systems can be used to develop a customizable dashboard-like reporting tool for educators that will extract and visualize real-time data on student engagement and likelihood of

success. Romero et al. [17] provided feedback to help decision making for improving student learning and taking the appropriate proactive action. Other examples and case studies for this category of applications can be found in [14].

- Domain structure analysis. The objective is to determine domain structure and improving domain models that characterize the content to be learned and optimal instructional sequences, using the ability to predict the student's performance as a quality measure of a domain structure model. Performance on tests or within a learning environment is utilized for this goal.
- Maintaining and improving courses. It is related to the two previous goals. The objective here is to determine how to improve courses (contents, activities, links, etc.), using information (in particular) about student usage and learning.
- Studying the effects of different kinds of pedagogical support that can be provided by learning software. For example, Anaya and Boticario [18] proposed a method to analyze collaboration using machine learning techniques.
- Advancing scientific knowledge about learning and learners through building, discovering or improving models of the student, the domain, and the pedagogical support. For example, Siemens and Baker [19] developed and tested a scientific theory about improving learning technology, and formulated a new scientific hypothesis.

We note that these EDM objectives aim to improve several aspects of educational systems in general and CBLE in particular. In this specific context, the learner modeling is a key point to accomplish several goals and tasks (tutoring, adaptation, personalization, etc.). Indeed, the different objectives depend heavily on this first objective "Student modeling" which is often supplemented by the behavior analysis, and therefore, allows the prediction of performance, generating recommendation, providing administrators and educators the adequate information to maintain and improve the content and learning environments.

Thus, if the EDM facilitates the modeling, and thus achieve the objectives mentioned above, several treatments become easier in CBLE. To accomplish these goals, educational data mining researches use the categories of technical methods described below.

1.2.4 The Used Methods

To achieve the EDM objectives, the majority of traditional data mining techniques including but not limited to classification, clustering, and association analysis techniques have been applied successfully in the educational domain. Nevertheless, educational systems have special characteristics that require a different treatment of the mining problem [14]. That is why researchers involved in EDM apply not only data mining techniques, but also propose, develop and apply methods and techniques drawn from the variety of areas related to EDM (statistics, machine learning, text mining, web log analysis, psychometrics, etc.).

The most popular classification of these methods is the one proposed in Baker [1]: prediction, clustering, relationship mining, distillation for human judgment and discovery with models. Bienkowski et al. [12] then Romero and Ventura [10] extended this taxonomy. Based on these studies and those in [20, 21] we regroup these techniques into the following methods:

- Prediction. The goal is to develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). Types of predictions methods are *classification* (when the predicted variable is a categorical value), *regression* (when the predicted variable is a continuous value), or *density estimation* (when the predicted value is a probability density function). An example of EDM application is predicting student's academic success [4] and behaviors [6].
- Clustering. Refers to finding instances that naturally group together and can be used to split a full dataset into categories. Typically, some kinds of distance measures are used to decide how similar instances are. Once a set of clusters has been determined, new instances can be classified by determining the closest cluster. In EDM, clustering can be used for grouping students based on their learning patterns or cognitive strategies [22].
- Relationship mining. Used for discovering relationships between variables in a dataset and encoding them as rules for later use. There are different types of relationship in mining techniques such as *association rule mining* (any relationships between variables), *sequential pattern mining* (temporal associations between variables), *correlation mining* (linear correlations between variables), and *causal data mining* (causal relationships between variables). In EDM, relationship mining is used to identify relationships between the students' online activities and the final marks [23] and to model learners' problem solving activity sequences [24].
- Distillation of data for human judgment. It is a technique that involves depicting data in a way that enables a human to quickly identify or classify features of the data. This approach uses summarization, visualization and interactive interfaces to highlight useful information and support decision-making. On the one hand, it is relatively easy to obtain descriptive statistics from educational data to obtain global data characteristics and summaries and reports on learner's behavior. On the other hand, information visualization and graphic techniques help to see, explore, and understand huge educational data at once. In [25] the visualization of sequences of student's activity helps to understand the patterns of learning environment use.
- Discovery with models. Its goal is to use a validated model of a phenomenon (using prediction, clustering, or knowledge engineering) as a component in further analysis such as prediction or relationship mining. It is used for example to identify the relationships between the student's behavior and characteristics [26].
- Outlier Detection. The goal of outlier detection is to discover data points that are significantly different than the rest of data. An outlier is a different observation (or measurement) that is usually larger or smaller than the other values in data.

In EDM, outlier detection can be used to detect deviations in the learner's or educator's actions or behaviors, irregular learning processes, and for detecting students with learning difficulties [27].

- Social Network Analysis. SNA or structural analysis, aims at studying relationships between individuals, instead of individual attributes or properties. SNA views social relationships in terms of network theory consisting of nodes (representing individual actors within the network) and connections or links (which represent relationships between the individuals, such as friendship, cooperative relations, etc.). In EDM, SNA can be used to interpret and analyze the structure and relations in collaborative tasks and interactions with communication tools [28].

- Process Mining. Its goal is to extract process related knowledge from event logs recorded by an information system to have a clear visual representation of the whole process. It consists of three subfields: conformance checking, model discovery, and model extension. In EDM, process mining can be used for reflecting students' behaviors in terms of their examination traces consisting of a sequence of course, grade, and timestamp triplets for each student [29].

- Text Mining. It is an extension of data mining to text that is focused on finding and extracting useful or interesting patterns, models, directions, trends, or rules from unstructured text documents such as HTML files, chat messages and emails. Text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling [10]. Text mining is used to analyze the content of discussion boards, forums, chats, Web pages, documents, etc. [3].

- Knowledge Tracing. KT is a popular method for estimating student mastery of skills that has been used in effective cognitive tutor systems. It uses both a cognitive model that maps a problem-solving item to the skills required, and logs of students' correct and incorrect answers as evidence of their knowledge on a particular skill. KT tracks student knowledge over time and it is parameterized by variables. There is an equivalent formulation of KT as a Bayesian network. In EDM, it is used for example for predicting student's behavior [30].

- Matrix Factorization. It is a decomposition of a matrix into a product of matrices. There are many matrix factorization techniques such as Non-negative Matrix Factorization (NMF). NMF consists of a matrix of positive numbers, as the product of two smaller matrices. For example, in the context of education, a matrix M that represents the observed examinee's test outcome data that can be decomposed into two matrices: Q that represents the Q-matrix of items and S that represents each student's mastery of skills [31]. Thai-Nghe et al. [32] used a matrix factorization model inspired from recommender systems to predict student performance.

We note here that an increasing number of techniques are used in EDM for the analysis of the different data produced in educational systems. The choice of which technique to use depends on the nature of the learning environment, the research objectives and the type of the available data. In what follows we discuss the type of the analyzed data.

1.2.5 The Analyzed Data

There are different analyzed data in EDM studies such as their objectives and techniques. We can distinguish these data according to the following features:

- Data availability:

  - Data already available recorded over the years in the institution databases (e.g. students' scores) or the log files of learning software.
  - Data generated during experiments within a research work.
  - Data available to researchers in benchmark repositories (PSL-Datashop[2], MULCE[3]).

- Collection sources:

  - Manual. Performed by a human observer that takes notes on the learning situation to evaluate the participants' activities.
  - Digital. Relies on the use of a hardware configuration that records the learner's activity. The result of such collection is a numerical trace that can be a log file, information stored in databases, audio or video records.
  - Mixed. Where both methods are used simultaneously.

- Learning environment [10]:

  - Traditional education. Primary, secondary, higher education, etc.
  - Computer-based education. Intelligent Tutoring System (ITS), Learning Management System (LMS), Adaptative Educational Hypermedia System (AEHS), Computer Supported Collaborative Learning (CSCL), serious games, test and quiz systems, etc.

- The educational described level [1, 9]:

  - The keystroke level, the answer level, the session level, the student level, the classroom level, the teacher level, and the school level.

- The type of data:

  - Qualitative or quantitative data.
  - Personal, administrative and/or demographic data (age, sex, etc.).
  - Answers to psychological questionnaires for measuring users' satisfaction, motivation, skills, cognitive features, etc.
  - Answers to questions and/or test scores of the academic system.
  - Individual interactions with the educational system: from fine grained actions such as mouse click, to high level ones such as number of attempts, the learner browsing pattern, etc.

---

2 https://pslcdatashop.web.cmu.edu/ visited on August 6, 2013.
3 http://mulce-pf.univ-fcomte.fr/PlateFormeMulce/visited on August 6, 2013.

– Social interaction (chat, sent messages, forum participation, etc.).
– Visual and facial reactions, etc.

We note here that the data are highly variable depending on the type of environment. In this chapter, we are interested in EDM applications on computer based education. In such systems, the collected data is often digital, and their size is often less important than traditional environments that have much bigger databases. However, several studies combine these two sources of data to give a complete view of the learner's behavior and performance. For instance, authors in [23] attempted to predict the success of students in the final exam based on their participation level in online forums. The fusion and the processing of these different types of data require several steps to implement the EDM process that we present in the following section.

1.2.6 Process of Applying the EDM

Romero and Ventura [10] and Sachin and Vijay [33] proposed a process of applying EDM close to the one of KDD (Knowledge discovery in databases) or other data mining application process (Fig. 1.2).

This process starts with collecting or choosing the data to study from the educational environment. The obtained raw data require cleaning and preprocessing (heterogeneous data fusion, treatment of missing and incorrect values, converting the data to an appropriate form, feature selection, etc.).

This phase often requires the use of some data mining techniques. That is why, and given its complexity some works try to eliminate this phase as [34] which provides a data model to structure data stored by Learning Management Systems,
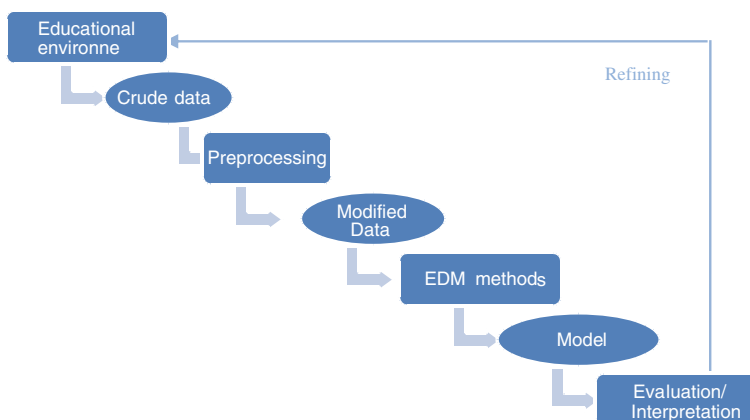

Fig. 1.2 The process of data mining application in educational data mining

and a tool that does the actual structure/export functionality, which they implemented for the Moodle LMS.

Once the data preprocessed, the appropriate EDM method/technique is applied. Finally, the last step is the interpretation and the assessment of the obtained results. To apply this process, which is often difficult given the heterogeneity of the data in the educational context, several tools are used.

1.2.7 Some Technological Tools Used in EDM

There are several tools and technologies used in the process of EDM not specifically designed for teaching and educational environments (Weka,[4] R,[5] etc.). However, in the last few years, a large number of data mining tools designed for educational purposes have been developed [10]. A summary of some of the most recent tools are presented in Table 1.1.

By analyzing these tools, we find that they are usually designed for computer-based educational systems. Moreover, apart from benchmark repositories (PSL-Datashop and MULCE), other tools are not re-used by other researchers of the EDM community.

This can be due to several reasons: their availability, the special format of the data to analyze, the difficulty of their deployment outside of their development environment or the ignorance of their existence. That is why; an effort should be made to make these tools available to the different learning actors (teachers, designers, administrators and researchers) to fulfill the different objectives and to analyze data from different environments.

Finally, now that we have an overview of the EDM field, we focus in the following on examples of its applications in computer-based educational systems. We particularly focus in behavior analysis and performance prediction and assessment.

1.3 Examples of EDM Applications in Computer-Based Learning Environments

In the last years, a wide number of EDM applications have been developed as seen in the previous sections. There are applications dealing with the assessment of students' learning performance, course adaptation and learning recommendations based on the student's learning behavior, evaluation of learning material and

---

[4] http://www.cs.waikato.ac.nz/ml/weka/ visited on August 6, 2013.

[5] http://www.r-project.org visited on August 6, 2013.

Table 1.1 Some tools for EDM applications

Tool	Goal	References
MMT tool	To facilitate the execution of all data mining steps in Moodle data	[2]
AAT	To analyze student's behavior in learning systems	[35]
E-learning web miner	To discover student's behavior profile in virtual course	[36]
SNAPP	To display the relation evolution between participants in discussion forums	[37]
EDM visualization tool	To visualize the process in which students solve procedural problems in logic	[38]
DRAL	To discover relevant e-activities for learners	[39]
IDLS	To analyze learner's browsing behavior and learning styles	[40]
eLAT	To explore and correlate learning object usage, user properties, user behavior, as well as assessment results based on graphical indicators	[41]
Meerkat-ED	To analyze interactions of students in asynchronous discussion forums of online courses	[28]
PSL–datashop	To store and analyze click-stream data, fine-grained longitudinal data generated by educational systems	[42]
MULCE		[43]

web-based courses, providing feedback to both teacher and students in e-learning courses, and detection of students' learning behaviors [21].

A review of these studies can be found in [9, 10, 12, 21, 44, 45]. Through these studies we noticed that the current mainstream EDM research is primarily focused on mining logs generated by the e-learning systems [13, 21]. We also found that the oldest and the most popular applications are the prediction of the student's performance and the analysis of learning behavior.

The term 'prediction' is generally used to characterize models (based on EDM techniques) designed for predicting new outcomes or scenarios based on new observations. Prediction is different from 'explanation', where the goal is to build models that explain underlying causal structure and to assess the explanatory power of such models [46]. This term is then linked to the study of the learner's behavior.

In the following, we present the most recent EDM studies from 2010 to 2013 related to these main objectives: learner's performance and behaviors in computer-based learning environments.

1.3.1 EDM Applications for Predicting and Evaluating Learning Performance

In this subsection, we analyze the current state of EDM research in learners' performance in CBLE. Table 1.2 summarizes some of the recent reviewed researches.

We note that the majority of the studied works applied EDM in LMS and ITS. In addition, as collaboration activities are often part of LMS, some studies treated collaborative data in LMS [23, 48, 56], while others [28, 55] analyzed collaboration usage data coming from a devoted environment.

The most tested LMS is Moodle. In [47] Jovanovica et al. applied classification models for predicting students' performance, and cluster models for grouping students based on their cognitive styles in Moodle. They developed a Moodle module that allows automatic extraction of data needed for educational data mining analysis and deploys models developed in this study. They indicate that the classification models helped teachers, students and business people, for early engaging with students who are likely to become excellent on a selected topic.

Furthermore, they indicate that clustering students based on cognitive styles and their overall performance enable better adaption of the learning materials with respect to their cognitive styles. Along the same lines, Falakmasir and Jafar [48] applied data mining methods (Feature Selection, decision trees) to the web usage records of students' activities in Moodle. As a result, they were able to identify and rank the students activities based on their impact on the performance of students in final exams/grades. Their findings suggest that students' participation in virtual classrooms had the greatest impact on their final grades.

Table 1.2 Some EDM applications for predicting the learner's performance in CBLE

References	CBLE	Data	Method	Goals
[47]	LMS: moodle	Data related to students' behaviors and cognitive style	Classification	To predict students' performance, and to cluster models for grouping students based on their cognitive styles
[48]	LMS: moodle	Web usage logs and final marks	Feature selection, classification (decision tree: C4.5 algorithm)	To identify and rank the students' activities based on their impact on the performance of students in final exams
[49]	LMS: moodle	Students' usage data in several courses	Statistical and classification methods (decision trees, rule and fuzzy rule induction methods, and neural network)	To predict the marks that university students will obtain in the final exam of a course
[23]	LMS: moodle	Students' usage data related to assignments and forums	Association rule mining algorithms	To evaluate the relation/influence between the on-line activities and the students' final marks
[46]	LMS: Sakai	Students' demographic data and Sakai log data of individual course events	Campbell's approach (factor analysis and logistic regression) classification, feature selection	To develop predictive models of students' success
[50]	Python tutor	Students' submissions and responses to satisfaction questionnaire	Statistics on the analyzed variables (means, average, etc.)	To generate personalized feedback and hints
[51]	Lecture capture system (LCS)	LCS log data combined with collected survey data	Statistics on the analyzed variables	To predict students' chance of passing exams
[32]	Computer-aided-tutoring systems	Log data of interactions (activities, success and progress indicators)	Matrix factorization	To predict student's performance
[52]	ITS: ASSISTment tutor	Students' responses	Bayesian networks	To predict student's performance by modeling skills at a fine-grained level

(continued)

Table 1.2 (continued)

References	CBLE	Data	Method	Goals
[30]	ITS: ASSiSTment tutor	Student's first response to a question in the tutor	Bayesian knowledge tracing, performance factors analysis, correct first attempt rate, straightforward averaging, regression AdaBoost, neural networks, random forest (random decision trees)	To improve students' performance prediction with ensemble methods
[53]	ITS: ASSiSTment tutor	Data related to test scores	Clustering (spectral clustering, k-means) bootstrap aggregation ensemble method	To improve students' performance prediction and to introduce the utility of using spectral clustering
[54]	ITS: algebra and bridge to algebra	Students' answers to questions	Classification (neural network, KNN) and matrix factorization	To predict a student's ability to answer questions correctly based on historic results
[31]	Intelligent learning environments	Real data (responses to question items) and simulated data generated by probability matrix	Non-negative matrix factorization (NMF), visualization techniques	To improve the mapping of skills to question items in order to assess the student's mastery level
[55]	Forums manuel	Forum participation	Classification via clustering	To predict the student's final marks
[28]	Computer-supported collaborative learning (CSCL)	Interactions in asynchronous discussion forums	Social network analysis	To assess students' collaborations and course participation and facilitate fairer evaluation of these participation in online courses
[56]	A web-based discussion board	Asynchronous web discussion	Statistical techniques (variable analysis: ANOVA test, a post hoc least significant difference (LSD) analysis)	To evaluate to what degree the different types of web-based discussion affected students' language production performance

Romero et al. [49] fulfill trials and demonstrated how web usage mining can be applied in the Moodle e-learning system to predict the marks that university students will obtain in the final exam of a course. They also identified several avenues for using classification in educational settings: discovering student groups with similar characteristics, identifying learners with low motivations, proposing remedial actions, predicting and classifying students using intelligent tutoring systems. In the same way authors in [23] studied student's usage data from a Moodle system related to quizzes, assignments and forum activities to evaluate the relation/influence between the on-line activities and the final mark obtained by the students. They used several association rule mining algorithms. The discovered rules predict students' exam results (fail or pass) based on their frequent activities and can also help the instructor to detect infrequent students' behaviors/activities. In [46] Lauria et al. used another LMS: Sakai. They used demographic data and the LMS log data of individual course events to develop a predictive model of student success. They used many EDM methods (factor analysis and logistic regression, C4/5/C5.0 decision trees, support vector machine (SVM) classifiers, Bayesian network) to build data mining models that can help predict students' performance and take corrective actions in higher education institutions.

Regarding ITS studies, Dominguez et al. [50] created a system to generate personalized feedback and hints by mining the student data collected by Python Tutor, an online learning system. They found that students who used the hinting system achieved significantly better results than those who did not, and stayed active on the site longer. Gorissen et al. [51] analyzed the interactions of students with the recorded lectures using educational data mining techniques. They found discrepancies as well as similarities between students' verbal reports and actual usage as logged by the recorded lecture servers. The data suggests that students who do this have a significantly higher chance of passing the exams [3]. Thai-Nghe et al. [32] analyzed students' interactions log files to build success and progress indicators in order to predict students' performance using matrix factorization.

In [30, 52, 53] the authors carried out several experiments using data related to test scores and students' responses on the ASSiSTment tutor. They applied many EDM methods (classification, clustering, Knowledge Tracing, etc.) to improve student's performance prediction. Toescher and Jahrer [54] analyzed students answering questions from two ITS: Algebra and Bridge to Algebra. They used a set of collaborative filtering techniques adopted from the field of recommender systems (ex. matrix factorization), to predict a student's ability to answer questions correctly, based on historic results. Similarly, Desmarais [31] used Non-negative Matrix Factorization on students' scores to determine the skills required for a given question, and how strong different students are for these skills.

Regarding collaboration, López et al. [55] used classification and clustering to predict students' final marks from their participation in forums. In the same way, Rabbany et al. [28] analyzed students' interactions in forum asynchronous discussion of online courses using Social Network Analysis to facilitate fairer evaluation of students' participation in online courses. They also proposed

Meerkat-ED, a specific, practical and interactive toolbox for analyzing students' interactions in asynchronous discussion forums.

In [56] Chang et al. used a web-based discussion board provided by an online educational platform to analyze students' language production. They used statistical techniques (ANOVA test, least significant difference (LSD) analysis) to evaluate to what degree the different types of web-based discussion affected students' language production performance.

Regarding the analyzed data, the majority of these studies used students' question responses since their general goal was the prediction of learners' success in the final exam, based on their responses to previous tests [31, 52, 54] or previous attempts [30]. However, some studies also used interaction traces [32, 48], communication traces [28, 56], and responses to satisfaction questionnaire [50] or combined between several types of data such as in [23, 44, 48, 51]. Another approach that we found in [31] was to generate simulated data using a probability matrix in order to test several models.

We also notice that the data set size is very variable: from 27 [56] to 4,927 [51] participants producing data over several hours, weeks, months or years attending a big data set size (over 20 millions in [54]). This is related to the context of the study and the data origin: data collected in experiments or dataset already available and used in previous experiments (e.g. the ASSiSTment tutor). This second alternative facilitates the analysis by avoiding the collection step, often not very obvious. Moreover, it even allows testing several methods and environment as in [30].

We also note that in these studies, the used tools are often not mentioned or are DM tools (Pentaho[6] in [46], RapidMiner[7], R in [31]) except for Rabbany et al. [28] who proposed their own tool (Meerkat-ED). Concerning the used method, clustering and prediction (classification) are on the top of the implemented techniques. However, several studies addressed the use of other techniques such as text mining, sequential pattern, SNA and matrix factorization. Statistical methods as well are used in many studies not only during the treatment phase of the EDM process but also during the preprocessing step where it is often difficult to choice the adequate features to use among the available data. For instance, in [46] "Feature Selection" is used to select the relevant attributes to use.

Through this study, we note that the application of different techniques of EDM allowed to identify the learner's performance (usually measured as the success of the learner in the final exam), from simple data (previous results or question answers, participation in collaborative activities, productions, etc.).

This can be exploited to improve the learning systems in different ways. For instance, if the majority of learners have low performance on a resource, it could hint to the fact that the course resource and/or the learning material are inadequate and therefore should be changed and/or improved.

---

[6] http://www.pentaho.com visited on August 6, 2013.

[7] http://rapid-i.com visited on August 6, 2013.

Some reviewed studies have discussed some of these results that contribute to a better adaptation and personalization of CBLE: improving adaptation based on cognitive styles in [47], classifying the learners' activities according to their influence on the performance in [48], identifying the required skills for a learning resource in [31].

Other studies also discussed results contributing to help the educators in their tutoring and assistance task: identifying learners with little motivation as in [49], grouping students based their characteristics in [48, 49], detecting infrequent behaviors in [23], etc. Thus we think that these results should be used to improve and adapt the content and the organization of the learning materials in CBLE, and could be used to the advantage of all the actors involved in the learning process.

1.3.2 EDM Applications for Analyzing Learners' Behaviors

In this subsection, we analyze the current state of EDM researches for analyzing (identifying, explaining, etc.) learners' behaviors in computer-based learning environments. Table 1.3 summarizes some of the recent reviewed research.

Among the reviewed studies we found three that belongs to LMS. Krüger et al. [34] aimed to build a data model to ease analysis and mining of educational data. To experiment their model, they analyzed the data stored in the "Programming 1" course in the Moodle LMS to study learners' behaviors related to solving self-evaluation exercises using association rule. They found that as the semester progresses, less students solve them. Macfadyen and Dawson [16] made an analysis of LMS racking data from a Blackboard Vista-supported course. The goal was to explain the variation in students' final grades. Using regression, they found significant correlation between the students' final grades and their learning behaviors on the LMS, based on key variables such as the total number of discussion messages posted, and the number of assessments completed.

In [40] Bousbia et al. aimed to automatically identify the learner's behavior and learning style, based on navigation trace analysis in a web-based learning environment: the eFAD LMS. They defined four browsing behaviors using a decision tree and carried out experiments using statistical techniques and machine learning classifiers (C4.5 decision tree, KNN, Bayesian networks, and neural networks).

Learner's behavior is also studied in other types of CBLE. Peckham and McCalla [22] carried out an experiment in a learning environment designed to emulate hypermedia courses to identify patterns of students' behaviors in a reading comprehension task using EDM techniques (k-means clustering, and ANOVA test).

Desmarais and Lemieux [25] also aimed to better understand the patterns of use of a learning environment. They applied clustering and activity sequence visualization on gathered logs of learners' interactions in a self-regulated web based drill and practice learning environment.

Table 1.3 Some EDM applications for analyzing learners' behaviors in CBLE

References CBLE	Data	Method	Goals
[34] LMS: moodle	Students' attempts for doing an exercise	Association rule	Analyze how students use learning resources in a course
[16] LMS: blackboard	Log data related to asynchronous discussion and assessments	Regression	Explain the variation in students' final grades
[40] LMS: eFAD	Log of navigation interaction	Classification: k-means, neural network, naïve bayes	Analyze learner's browsing behaviors styles
[22] Learning environment designed to emulate hypermedia courses	Log data (events, actions)	K-means clustering, ANOVA	Identify students' behavior patterns in a reading comprehension task
[25] Self-regulated web based drill and practice environment	Logs of learners' interactions (answers to exercises, actions, events)	Clustering, classification, activity sequence, visualization	Understand the patterns of use of a learning environment
[3] A live video streaming environment	Students' online interaction	Text mining	Examine students' online interaction
[57] AutoTutor a dialogue tutor. The incredible machine TIM: a problem solving game), and Aplusix. A problem-solving based ITS	Study 1: standard pre-test intervention post-test, video of participants and of the content of their screen. Study 2 and 3: students' observation made by observers	Human judgments about the same set of six cognitive affective states, ANOVA test	Study the incidence, persistence, and impact of students' cognitive-affective states
[26] MetaTutor, an agent-based ITS	Trace data about tests	Clustering: expectation–maximization, sequence mining	Identify distinguishing patterns of behaviors
[58] Betty's brain: learning-by-teaching environment	Students' learning interactions as they teach an agent.	Sequence mining techniques, linear segmentation algorithm	Evaluate and compare students' learning behaviors

In [3] a live video streaming (LVS) system was used to study the students' patterns using data mining and text mining applied on data of online interaction. Bouchet et al. [26] analyzed students' characteristics and learning behaviors in MetaTutor, an agent-based ITS. They used clustering and sequence mining to distinguish patterns of behaviors. Similarly, Kinnebrew and Biswas [58] used sequence mining to identify learning behaviors in Betty's Brain, a learning-by-teaching environment.

In [57] Baker et al. carried out three studies in three CBLE: AutoTutor (a dialogue tutor), the incredible machine (TIM) (a problem solving game: a simulation environment), and Aplusix (a problem-solving based ITS). The studied data were pre-test–intervention–post-test, and video records of the participants and their computer screen in the first study, and observation made by observers related to cognitive affective states on the second and third studies. Using Human judgment and ANOVA test, the authors found that boredom was very persistent across learning environments and was associated with poorer learning and problem behaviors, such as gaming the system. Also, confusion and engaged concentration were the most common states within the three learning environments. These findings suggest that significant effort should be put into detecting and responding to boredom and confusion.

Throughout Table 1.3, we notice that all the reviewed studies used interaction traces, which are generally of low level (action/event) or specific to the analyzed activity (messages, reading task, etc.). These dataset are often structured in numerical attributes were task scores or statistics on log data (frequencies of actions, time spent in actions) are the most used. Moreover, the sample size used in these studies is less variable since almost all these works are based on experiments (from 28 in [22] to 148 participants in [26]). We note that the used tools for analysis are often not mentioned in these studies. The two mentioned ones are Weka in [26] and TraMiner-R in [25].

Regarding the used methods, clustering and classification still on the top as the majority of the presented works aim to identify common learning behaviors. Other methods were also used such as text mining, sequence mining, statistical methods (e.g. to calculate some variables) as well as Human judgments when the analyzed data referred to personal characteristics.

Thus, through the reviewed studies presented here, we find that EDM allows from low level traces to analyze and evaluate the student's behavior. This task is often a difficult one given the close relationships of the behavior to personal characteristics such as learning styles, emotions and its frequent changes according to the learner's state, the learning time, the type and the content of the learning materials, the learner's reaction to other actors, etc. Thus, behavioral analysis should be done in real time to provide a better feedback to teachers as well as learners in order to improve the tutoring and learning tracking tasks. This is still difficult even with the use of EDM techniques regarding the small size of the analyzed samples which does not allow the generalization of the obtained results that remain specific to the studied environments and the context of the carried out experiments.

However, even if the majority of research, such as those presented here, focus on the analysis of the past behaviors to explain a phenomena such as abandon, or evaluate the participation and the obtained results, their findings should be used to improve learning environments based on the students' behavioral patterns.

1.3.3 Discussion

The 25 reviewed studies presented in this section give an overview of the typical educational environment, data and methods used in EDM applications. LMS or generally online educational environment and ITS are the most exploited. This is probably due to their wide use in the educational environment, which facilitates the realization of experiments, which is often the data collection source of these studies.

The analyzed data are generally related to assessments (tests, quizzes, exams, etc.), fine grained online interaction (action, event) and also participation in collaboration activities. This data type is related to the type of the studied CBLE that provide such information in their database and log files, as well as the sighted objectives of these studies. However, some researches combined these data with the video recording of learners or the human manual observations during the learning sessions. This combination, although difficult to achieve, can refine the study especially in CBLE where there is less face to face interactions between the teacher and the learner.

Regarding EDM methods, the most used were prediction (classification, association rules, and regression) and clustering. This finding can be explained by the two objectives studied: learner's performance and behavior, and by the fact that these techniques are mature, widely known, tested and implemented in the DM used tools, and also provide satisfactory results even with small sample size, we often find here.

Other methods were also used, according to the analyzed data and the objective to exploit other techniques and improve the results. Note, however, that it is not easy to identify for a given CBLE type, a given type and size of data set, and a given goal, which is the best EDM technique to use. Certainly this information helps to establish a choice, but it does not limit or confirm that this is the best one.

This observation explains why in several studies several techniques were used to achieve the best results. Note that we did not discuss in this chapter the percentage of the obtained results, since they depend on the different context of the studied works (types of environment and the analyzed data, the student populations, the set parameters and hypothesis during the EDM process, etc.). Indeed, although the obtained results are generally satisfactory in their context, they remain is an experimental stage and cannot be generalized.

Finally, we note that although we focused on the study of EDM applications related to two main objectives, namely the prediction of performance and behavior analysis, the results of the presented research achieved other EDM objectives

(student modeling, communicating to stakeholders, maintaining and improving courses, etc.). In addition, as the two studied objectives are closely related, we found studies dealing them both where EDM techniques were applied to explore the relationships between the learner's behavior and the learning performance, to improve the learning environment.

For example, in [59] learners' behaviors is used to predict the success or the failure of students without requiring the results of formal assessments. In [60] Bayer et al. focused on predicting drop-outs and school failures when students' data have been enriched with data derived from students' social behaviors.

We think that this last objective of analyzing the reasons of failure, drop-out and abandon is a promoter research area in the EDM field that should to be exploited in the future, especially for CBLE, where it is a common phenomenon. We also believe that the data collected from these environments should be enriched by other types of information such as demographic data, to provide a better explanation of the observed phenomena. An effort should also be provided to share the analyzed samples and provide significant benchmark to pass the experimental stage in order to generalize the established models and the results found in these studies to improve the learning environments.

For the same goal, it is required to improve EDM tools. In fact, although DM tools allow the analysis, they require some expertise to set the parameters and make the appropriate interpretation. It is therefore necessary that EDM have their own tools to make these techniques within the reach of teachers, and allow more advanced treatment combining multiple data sources, and proposing some methods according to the type of these data and the analyze goal. So we can imagine these tools included in learning environments to facilitate their access to the different learning actors.

1.4 Conclusions

In this chapter, we discussed the use of EDM in educational systems. We studied recent EDM applications (2010–2013) by taking into account: the educational system, the analyzed data, the used method for the analysis, the used tool, and the analysis goal, especially in computer based learning environments.

We noticed through this study that a large number of researches are interested today in the application of EDM in educational systems in general and in CBLE in particular, to exploit the available data or the one that can be collected in these environments to ensure their improvement through the various objectives. We can say that EDM introduces a major advantage, drawn from data mining and KDD fields, the one related to extract hidden information about learners and learning from recorded data.

We have reviewed, in some detail, recent research dealing with students' performance and behaviors. We found that the use of EDM methods helps the prediction of students' performance; especially final marks. It also helps to identify

and explain usual and unusual learning behaviors that should facilitate the assistance of learners, and reduce the costs of educative personalization and adaptation processes.

However, these contributions have to go out of laboratories to be applied in the used educational systems in order to improve learning. Studies in this goal are initiated especially in traditional educational environments where the results of the application of EDM on existent data are used to improve the educational system [5].

We expect however, that this goal will be also applied in software educational systems, to find new ways to improve learning materials and reduce the abandon rate that is considerable in such environments. In fact, to make this area more mature, it is necessary that the established models in these studies could be tested in real environments for frequent use to affirm and exploit the found results to improve these environments.

A first step in this direction is the sharing and the reuse of the dataset through open data repositories and standard data formats to promote the exchange of data and models. It is also necessary to popularize the use of EDM through the popularization of tools targeted to the different learning actors for the analysis of educational data in a simple and intuitive way, while providing suggestion about methods to apply for a better result and facilitating the interpretation of these results. We think that it is also necessary to take into account the EDM process in the overall development process of the computer based learning environment to ensure a significant improvement.

References

1. Baker, R.S.J.d.: Data mining for education. In: McGaw, B., Peterson, P., Baker, E. (eds.) International Encyclopedia of Education, vol. 7, 3rd edn., pp. 112–118. Elsevier, Amsterdam (2010)
2. Pedraza-Perez, R., Romero, C., Ventura, S.: A java desktop tool for mining moodle data. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero, C., Stamper, J. (eds.) Proceedings of 4th International Conference on Educational Data Mining, pp. 319–320. International Educational Data Mining Society, Eindhoven (2011)
3. He, W.: Examining students' online interaction in a live video streaming environment using data mining and text mining. Comput. Hum. Behav. **29**(1), 90–102 (2013)
4. Ayesha, S., Mustafa, T., Sattar, A., Khan, I.: Data mining model for higher education system. Eur. J. Sci. Res. **43**(1), 24–29 (2010)
5. Pal, S.: Mining educational data to reduce dropout rates of engineering students. Int. J. Inf. Eng. Electron. Bus. **2**(1), 1–7 (2012)
6. Parack, S., Zahid, Z., Merchant, F.: Application of data mining in educational databases for predicting academic trends and patterns. In: Proceedings of 2012 IEEE International Conference on Technology Enhanced Education, pp. 1–4. IEEE Press, Piscataway (2012)
7. Huebner, R.A.: A survey of educational data-mining research. Res. High. Educ. J. **19**, 1–13 (2013)
8. Calders, T., Pechenizkiy, M.: Introduction to the special section on educational data mining. ACM SIGKDD Explor. **13**(2), 3–6 (2011)

9. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **40**(6), 601–618 (2010)
10. Romero, C., Ventura, S.: Data mining in education. Wiley Interdisc. Rev.: Data Min. Knowl. Discovery **3**(1), 12–27 (2013)
11. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. Int. J. Technol. Enhanced Learn. **4**(5–6), 318–331 (2012)
12. Bienkowski, M., Feng, M., Means, B.: Enhancing teaching and learning through educational data mining and learning analytics: an issue brief. US Department of Education, Office of Educational Technology, pp. 1–57 (2012)
13. Scheuer, O., McLaren, B.M.: Educational data mining. In: Seel, N.M. (eds.) Encyclopedia of the Sciences of Learning, pp. 1075–1079. Springer, US (2012)
14. Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d.: Introduction. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (eds.) Handbook of Educational Data Mining, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series, pp. 1–5. CRC Press, Boca Raton (2011)
15. Kotsiantis, S., Patriarcheas, K., Xenos, M.: A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowl.-Based Syst. **23**(6), 529–535 (2010)
16. Macfayden, L.P., Dawson, S.: Mining LMS data to develop an ''early warning'' system for educators: a proof of concept. Comput. Educ. **54**(2), 588–599 (2010)
17. Romero, C., Zafra, A., Luna, J.M., Ventura, S.: Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. Expert Syst. **30**(2), 162–172 (2013)
18. Anaya, A.R., Boticario, J.G.: Application of machine learning techniques to analyse student interactions and improve the collaboration process. Expert Syst. Appl. **38**, 1171–1181 (2011)
19. Siemens, G., Baker, R.S.J.d.: Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of 2nd International Conference on Learning Analytics and Knowledge, pp. 1–3. ACM, New York (2012)
20. Baker, R.J.D.F., Yacef, K.: The state of educational data mining in 2009: a review and future visions. J. Educ. Data Min. **1**(1), 3–17 (2009)
21. ALMazroui, Y.A.: A survey of data mining in the context of e-Learning. Int. J. Inf. Technol. Comput. Sci. **7**(3), 8–18 (2013)
22. Peckham, T., McCalla, G.: Mining student behavior patterns in reading comprehension tasks. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 87–94. International Educational Data Mining Society, Chania (2012)
23. Romero, C., Romero, J.R., Luna, J.M., Ventura, S.: Mining rare association rules from e-learning data. In: Baker, R.S.J.D., Merceron, A., Pavlik Jr., P.I. (eds.) Proceedings of 3rd International Conference on Educational Data Mining, pp. 171–180. International Educational Data Mining Society, Pittsburgh (2010)
24. Kock, M., Paramythis, A.: Activity sequence modeling and dynamic clustering for personalized e-learning. User Model. User-Adap. Inter. **21**(1–2), 51–97 (2011)
25. Desmarais, M.C., Lemieux, F.: Clustering and visualizing study state sequences. In: D'Mello, S.K., Calvo, R.A., Olney, A. (eds.) Proceedings of 6th International Conference on Educational Data Mining, pp. 224–227. International Educational Data Mining Society, Memphis (2013)
26. Bouchet, F., Azevedo, R., Kinnebrew, J.S., Biswas, G.: Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self regulated learning. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 65–72. International Educational Data Mining Society, Chania (2012)
27. Barahate. S.R.: Educational data mining as a trend of data mining in educational system. In: Proceedings of IJCA International Conference and Workshop on Emerging Trends in Technology, pp. 11–16 (2012)

28. Rabbany, R., Takaffoli, M., Zaïane, O.: Analyzing participation of students in online courses using social network analysis technique. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero, C., Stamper, J. (eds.) Proceedings of 4th International Conference on Educational Data Mining, pp. 21–30. International Educational Data Mining Society, Eindhoven (2011)
29. Trčka, N., Pechenizkiy, M., Aalst W.v.d.: Process mining from educational data. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (eds.) Proceedings of Handbook of Educational Data Mining, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series, pp. 123–142. CRC Press, Boca Raton (2011)
30. Pardos, Z.A., Gowda, S.M., Baker, R.S.J.d., Heffernan, N.T.: The sum is greater than the parts: ensembling models of student knowledge in educational software. ACM SIGKDD Explor. **13**(2), 37–44 (2011)
31. Desmarais, M.C.: Mapping question items to skills with non-negative matrix factorization. ACM SIGKDD Explor. **13**(2), 30–36 (2011)
32. Thai-Nghe, N., Drumond, L., Krohn -Grimberghe, A., Schmidt-Thieme, L.: Recommender system for predicting student performance. Procedia Comput. Science **1**(2), 2811–2819 (2010)
33. Sachin, B.R., Vijay, S.M.: A survey and future vision of data mining in educational field. In: Proceedings of IEEE 2nd International Conference on Advanced Computing and Communication Technologies, pp. 96–100. ACM, New York (2012)
34. Krüger, A., Merceron, A., Wolf, B.: A data model to ease analysis and mining of educational data. In: Baker, R.S.J.D., Merceron, A., Pavlik Jr., P.I. (eds.) Proceedings of 3rd International Conference on Educational Data Mining, pp. 131–140. International Educational Data Mining Society, Pittsburgh (2010)
35. Graf, S., Ives, C., Rahman, N., Ferri, A.: AAT: a tool for accessing and analysing students' behaviour data in learning systems. In: Proceedings of 1st International Conference on Learning Analytics and Knowledge, pp. 174–179. ACM, New York (2011)
36. Zorrilla, M., Garcia-Saiz, D.: A service oriented architecture to provide data mining services for non-expert data miners. Decis. Support Syst. J. **55**(1), 399–411 (2013)
37. Bakharia, A., Dawson, S.: SNAPP: a bird's-eye view of temporal participant interaction. In: Proceedings of 1st International Conference on Learning Analytics and Knowledge, pp. 168–173. ACM, New York (2011)
38. Johnson, M., Barnes, T.: EDM visualization tool: watching students learn. In: Baker, R.S.J.D., Merceron, A., Pavlik Jr., P.I. (eds.) Proceedings of 3rd International Conference on Educational Data Mining, pp. 297–298. International EDM Society, Pittsburgh (2010)
39. Zafra, A., Romero, C., Ventura, S.: DRAL: a tool for discovering relevant e-activities for learners. Knowl. Inf. Syst. **36**(1), 211–250 (2013)
40. Bousbia, N., Rebaï, I., Labat, J.-M., Balla, A.: Learners' navigation behavior identification based on traces analysis. User Model. User-Adap. Inter. **20**(5), 455–494 (2010)
41. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U.: Design and implementation of a learning analytics toolkit for teachers. Educ. Technol. Soc. **15**(3), 58–76 (2012)
42. Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A data repository for the EDM community: the PSLC datashop. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (eds.) Proceedings of Handbook of Educational Data Mining, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series, pp. 43–55. CRC Press, Boca Raton (2011)
43. Reffay, C., Betbeder, M.-L., Chanier, T.: Multimodal learning and teaching corpora exchange: lessons learned in 5 years by the Mulce project. In: special issue on dataTEL: datasets and data supported learning in technology-enhanced learning. Int. J. Technol. Enhanced Learn. **4**(1–2), 11–30 (2012)
44. Kotsiantis, S.B.: Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif. Intell. Rev. **37**(4), 331–344 (2012)

45. Amershi, S., Conati, C.: Combining unsupervised and supervised classification to build user models for exploratory learning environments. J. Educ. Data Min. **1**(1), 18–71 (2009)
46. Lauria, E., Baron, J.: Mining Sakai to measure student performance: opportunities and challenges in academic. In: Proceedings of Enterprise Computing Community Conference (2011)
47. Jovanovica, M., Vukicevica, M., Milovanovica, M., Minovica, M.: Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. Int. J. Comput. Intell. Syst. **5**(3), 597–610 (2012)
48. Falakmasir, M., Jafar, H.: Using educational data mining methods to study the impact of virtual classroom in e-learning. In: Baker, R.S.J.D., Merceron, A., Pavlik Jr., P.I. (eds.) Proceedings of 3rd International Conference on Educational Data Mining, pp. 241–248. International Educational Data Mining Society, Pittsburgh (2010)
49. Romero, C., Espejo, P.G., Zafra, A., Romero, J.R., Ventura, S.: Web usage mining for predicting final marks of students that use moodle courses. Comput. Appl. Eng. Educ. J. **21**(1), 135–146 (2013)
50. Dominguez, A.K., Yacef, K., Curran, J.: Data mining to generate individualised feedback. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010, Part II. LNCS, vol. 6095, pp. 303–305. Springer, Heidelberg (2010)
51. Gorissen, P., Bruggen, J., Jochems, W.: Usage reporting on recorded lectures using educational data mining. Int. J. Learn. Technol. **7**(1), 23–40 (2012)
52. Pardos, Z.A., Heffernan, N.T., Anderson, B.S., Heffernan, C.L.: Using fine-grained skill models to fit student performance with Bayesian networks. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (eds.) Handbook of Educational Data Mining, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series, pp. 417–426. CRC Press, Boca Raton (2011)
53. Trivedi, S., Pardos, Z.A., Sárközy, G.N., Heffernan, N.T.: Spectral clustering in educational data mining. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero, C., Stamper, J. (eds.) Proceedings of 4th International Conference on Educational Data Mining, pp. 129–138. International Educational Data Mining Society, Eindhoven (2011)
54. Toescher, A., Jahrer, M.: Collaborative filtering applied to educational data mining. J. Mach. Learn. Res. (2010)
55. López, M.I., Luna, J.M., Romero, C., Ventura, S.: Classification via clustering for predicting final marks based on student participation in forums. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 148–151. International EDM Society, Chania (2012)
56. Chang, M.M., Lin, M.C., Tsai, M.J.: A study of enhanced structured web-based discussion in a foreign language learning class. Comput. Educ. **61**, 232–241 (2013)
57. Baker, R.S.Jd., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. Int. J. Hum.-Comput. Stud. **68**(4), 223–241 (2010)
58. Kinnebrew, J.S., Biswas, G.: Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 57–64. International EDM Society, Chania (2012)
59. McCuaig, J., Baldwin, J.: Identifying successful learners from interaction behaviour. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 160–163. International Educational Data Mining Society, Chania (2012)
60. Bayer, J., Bydzovska, H., Geryk, J., Obsıvac, T., Popelınsky, L.: Predicting dropout from social behaviour of students. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., Stamper, J. (eds.) Proceedings of 5th International Conference on Educational Data Mining, pp. 103–109. International Educational Data Mining Society, Chania (2012)