

# Application of Data Mining and Process Mining approaches for improving e-Learning Processes

K. Grigorova<sup>a</sup>, E. Malysheva<sup>b</sup>, S. Bobrovskiy<sup>b</sup>

<sup>a</sup> Angel Kanchev University of Ruse, 7017, 8 Studentska street, Ruse, Bulgaria

<sup>b</sup> Volga Region State University of Services, 445677, 4 Gagarina street, Togliatti, Russia

---

## Abstract

The paper describes the basic principles of Data mining and Process mining methods, their similarities and differences. Some examples of special software for Data mining and Process mining are presented. The authors examine the research in Educational Data Mining field, associated with the use of Data mining techniques in education, give examples of problems to be solved with the use of Data mining and Process mining techniques in the area of traditional and e-learning, describe the possibilities and limitations of different methods. A review of major scientific conferences and journals devoted to the research in Educational Data Mining is made.

*Keywords:* data mining; process mining; education data mining; e-Learning

---

## 1. Introduction

Currently, modern information systems have accumulated a huge amount of data about processes taking place in the various domain areas. These systems comprise ERP, CRM, SCM, WFM, B2B, B2C, and many others, including e-Learning system. Many of today's information systems collect and store data about the events occurring during the systems' performance in so-called event logs. Data mining and Process mining technologies allow the use of the event log data for analysis and improvement of the processes. Availability of advanced software dealing with Data mining and Process mining, allows to test these techniques on data obtained from real processes. A stimulus for the growing interest in Data mining and Process mining is the constant increase in the amount of data recorded in the information systems, including data about events that provide detailed information about the history of the processes, and the need to improve and support business processes in competitive and rapidly changing environment.

## 2. Data Mining and Process Mining: An Overview

Data mining - a multidisciplinary area, which has arisen and developed on the basis of such science fields as applied statistics, artificial intelligence, pattern recognition, machine learning, algorithmization, database theory and others. Here are some definitions of the concept. Gartner Group, the agency that analyzes the information technology markets, defines Data mining as follows: "The process of discovering meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories. Data mining employs pattern recognition technologies, as well as statistical and mathematical techniques" [1]. SAS Institute, a developer of analytical software, mentions in his definition of big data and its practical usefulness: "Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more" [2]. In the Data mining Curriculum [3] the following definition is met: "Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems". Data mining might consist of the following steps: identification of patterns and associations (free search), the use of the association rules to predict unknown values (predictive analytics), identification and analysis of the exceptions in the identified rules (anomaly detection).

Data mining methods and algorithms include: artificial neural networks, decision trees, symbolic rules, nearest neighbor method, support vector machines, Bayesian networks, linear regression, correlation and regression analysis, cluster analysis, association rules support, programming and genetic algorithms, a variety of methods for data visualization and many others. Most of the analytical methods used in Data mining technology are well-known mathematical algorithms and methods. New in their application is the possibility to use them in solving various concrete problems, due to existing appropriate hardware and software.

Process mining is a relatively young research discipline that sits between Big data and Data mining on the one hand, and Business Process Modeling and Analysis on the other. The idea of Process mining is to detect, control and improve the actual occurring processes by extracting knowledge from event logs readily available in modern information systems [4], [5]. Process mining includes automated process detection (extraction the process models from event logs), conformance checking (monitoring deviations by comparing model and event logs), defining the organizational structure, automated construction of simulation models, model extension and recovery, the prediction of process behavior in order to develop recommendations on the basis of the process history.

Process mining techniques provide new means for detecting, monitoring and improvement of processes in various fields of application, offer opportunities for a stricter conformance checking and the validation and reliability of information about the

basic processes of the organization. It is an important tool for modern organizations that need to manage non-trivial operational processes, since on the one hand, there is an incredible growth of event data, on the other hand, the processes should be aligned with the need for effective customer service.

At the core of both methods (Process mining and Data mining) are the data. They have a lot in common, as they use the same mathematical algorithms and techniques. The main difference is that Data mining operates with the data in general, whilst Process mining works with the data about events, which contain information about the processes [4]. Data mining and Process mining are complementary approaches that can reinforce each other. Process models detected and aligned with the event log data confirm the value of data analysis and provide a basis for further development as of Process mining, as well as of Data mining.

Special software is necessary for the implementation of Data mining and Process mining. More and more software vendors add to their software products such features. Examples of software products with the capabilities of Data mining and Process mining are presented in Table 1.

**Table 1.** Examples of software products with the capabilities of Data mining and Process mining

Tool Name	Vendor	Website
ProM	Open Source	www.promtools.org
ProM Lite	Open Source	www.promtools.org
RapidProM	Open Source	www.rapidprom.org
RapidMiner	Open Source	www.rapidminer.com
WEKA	Open Source	www.cs.waikato.ac.nz/ml/weka/
Orange	Open Source	orange.biolab.si
NLTK	Open Source	www.nltk.org
Celonis Process Mining	Celonis GmbH	www.celonis.de
Disco	Fluxicon	www.fluxicon.com
Minit	Gradient ECM	www.minitlabs.com
Perceptive Process Mining	Lexmark	www.lexmark.com
QPR ProcessAnalyzer	QPR	www.qpr.com
Rialto Process	Exeura	www.exeura.eu
SNP Business Process Analysis	SNP AG	www.snp-bpa.com
SPSS	IBM	www-01.ibm.com/software

One of the commonly used software is freeware ProM. ProM has over 1,500 plug-ins, allowing the use of different methods and algorithms for Data mining and Process mining, different types of data and models, to convert the data and models, etc., and the version ProM Lite contains the most commonly used modules. Most commercial software products, including Data mining and Process mining, are easy to use. Approximately 40 software products, often used in Data mining in the field of education are given in [6].

### 3. Data Mining and Process Mining in Education and e-Learning

One of the main directions of modern Data mining application is Educational Data Mining (EDM). The main goal of EDM is to use the huge amount of data about the educational processes, coming from different sources in different formats and with different levels of detail. The data represents information about the educational process, provides better understanding of learning and improving its outcomes.

#### 3.1. Data and problems in EDM

During EDM researchers use data of educational systems such as intelligent computer-based training, distance learning systems, online classes, and discussion forums, electronic manuals, school information systems, computer-aided testing system [7].

EDM analyzes the data by any type of information system, supporting training or education (schools, colleges, universities and other academic or professional education institutions, providing traditional and modern forms and methods of training, and informal learning). These data are not limited to the interaction of individual students with the educational system (e.g. navigating through the training and testing system, data entry in the tests and interactive exercises), but may also include data about the cooperation of students (for example, text chat), administrative data (e.g. school, district, teacher), demographics (e.g. gender, age, school classes), student emotionality (e.g. motivation, emotional state) and so on. The data have typical characteristics, such as multiple levels of hierarchy (a level for subject, a level for grading, a level for question), the context (a

specific student in a particular class answers to a specific question in a particular time on a particular date), short time data (recording data with different resolutions to facilitate various analyses, for example, to record data every 20 seconds) and long periods of time data (a big amount of data recorded over many sessions over an extended period of time, for example, covering semester and yearly courses) [8].

Nowadays in the field of education there are a wide variety of educational environments and information systems. Computer-based education (CBE) refers to the use of computers in education to provide directed training to generate control instructions for the student. The first CBE systems are a stand-alone educational applications that work on your computer without the use of artificial intelligence for student modeling, adaptation, personalization, and so on. Global use of the Internet has led to development of many new Web based educational system, such as e-learning systems, distance learning systems, on-line training systems, and so on, and the increasing use of artificial intelligence has led to the emergence of new intelligent and adaptive educational systems. The main types of currently used systems include: Learning management systems (LMS) [9], Intelligent tutoring systems (ITS) [10], adaptive intelligent hypermedia systems (AIHS) [11], Test and quiz systems [12] and others. Each of them provides a variety of data sources that need to be processed in different ways depending on the nature of the available data and the specific problems and tasks that are solved by using Data mining techniques.

EDM is often used when working with educational programs, in solving problems of modeling student's behavior and forecasting of the course results. Since the main purpose of Data mining in the field of education is to greatly improve the quality of training, it is more difficult to get quantitative measurements than in other areas, and the results should be evaluated through indicators like improving efficiency. Thus, a data-driven decisions are formed aiming to improve the current educational processes and teaching materials.

Examples of problems solved with the help of EDM, are:

- Predicting student achievement, assessment of knowledge and learning outcomes [12];
- Formation of recommendations to students based on their interests and activities in the learning process [13];
- Individual approach, adapting training to each student, including course content, navigation on the course, the presentation of the material [14], [15];
- Provision feedback to the teachers in order to support decision-making on the improvement of student learning, the adoption of pre-emptive actions to remedy the situation [12];
- Development and validation of scientific theories on learning technology, the formation of new scientific hypotheses [16];
- Monitoring the progress of learning to detect in real time the undesirable behavior of students, such as the termination of training, low motivation, incorrect use of educational forums, abuse, fraud, etc., creating warnings to the parties concerned [17];
- Cognitive models of students presenting their skills and knowledge (User / Student modeling) [18];
- Simulation the domain teaching instructions in terms of concepts, skills, training modules and their relationships [19];
- Identification the groups of students according to their individual characteristics, personal characteristics, features of the training, etc. [20] [21];
- Building a curriculum and educational content [22], [23];
- Planning and scheduling of future courses, course planning, planning of resource allocation, organization of access to learning materials, planning consultations, curriculum development, etc. [24];
- Estimation of parameters of probability models based on data about learning to determine the likelihood of events of interest [25].

A variety of problems and their educational performance leads to the need to adapt methods of Data mining and Process mining to these data and problems. The applicability of Data mining techniques in the field of education are considered in [8], [26].

### *3.2. Data Mining and Process Mining methods in e-Learning systems*

In EDM, the most commonly used methods are classification, clustering, Bayesian modeling, relationship mining and discovery with models, as well as the detection of anomalies, Social network analysis, Text mining, Distillation of data for human judgment, Knowledge tracing, nonnegative Matrix factorization and techniques and algorithms of Process mining.

Prediction – a definition of how the target attribute depends on a combination of other attributes. The types of forecasting methods are: classification (target variable is a category), regression (target and background variables are numbers), the density score (predicted value is the probability density function). Using these methods to predict student performance and to determine the pattern of student behavior is considered in [27] and [28].

Clustering is identification of groups of similar instances. Typically, to determine the similarity the distance measure is used. After the set of clusters is determined, new items can be classified according to the nearest cluster. The clustering in EDM can be used to group similar course materials or to form groups of students based on their knowledge and patterns of interactions [29], [30]. Examples of the applicability of various types of clustering algorithms in EDM are discussed in [31].

Outlier Detection - is to identify the data that are significantly different of rest of the data. Abnormal values correspond to the observations (or measurements), which are usually more or less than other values. The EDM anomaly detection can be used for the detection of students with learning difficulties, deviations in the actions or behavior of a student or a teacher, and for the detection of irregular learning processes [32].

Relationship Mining is to determine the relationships between the variables and presenting them in the form of rules for subsequent use. There are different types of relationship mining, such as association rule mining (relations between variables), sequential pattern mining (temporal association between variables), correlation mining (linear correlation between variables) and causal data mining (the causal relationships between variables). In EDM, relationship mining can be used to determine the relationships in student behaviors (behavior patterns) and to diagnose difficulties in teaching or the mistakes that often occur together. [33]

Social Network Analysis (SNA) allows us to understand and to measure the relationship between the entities in the network information. SNA considers social relationships in terms of network theory consisting of nodes (representing individual actors within the network) and the connections or ties (which represent relationships between individuals, such as friendship, kinship, organizational position, etc.). In the EDM SNA can be used to obtain information to interpret and analyze the structure and relationships in the interaction tasks, including interaction with the communications [34].

Text Mining, also known as text data mining and text analytics, is a method of producing high-quality information from text. Typical tasks include text mining categorization of text, text clustering, concept / entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling. In the EDM, text mining was used to analyze the content of discussion boards, forums, chats, Web pages, documents, and so on. [35].

Distillation of Data for Human Judgment is to present the data in an understandable form using generalization, visualization and interactive interfaces to extract useful information and to support decision making. This method comprises obtaining statistical data about the learning process to determine the common characteristics, obtaining summary data and reports on the behavior of the trainee. Data visualization and graphical techniques help to see, explore and understand large amounts of educational data immediately. In the EDM is also known as distillation for human judgment [36] and it has been used to assist teachers with the visualization and analysis of the students activity and the use of the information. [37]

Discovery with Models is to use previously tested phenomena model (using a prediction, clustering, or manual knowledge engineering) as a component of another kind of analysis such as prediction or relationship mining [38]. This method is often used in EDM and supports the identification of the relationship between the student's behavior and its characteristics, the use of psychometric modeling systems in machine-learning models, the analysis of research in various fields of study [39].

Knowledge Tracing (KT) is a popular method to assess student skills, which is used in effective cognitive tutor systems [40]. It uses a cognitive model that maps problem-solving item required skills and records correct and incorrect responses of students as evidence of their knowledge of a particular skill. KT monitors students' knowledge for some time, and parameterizes them by four variables. KT corresponds to the method of Bayesian network.

Nonnegative Matrix Factorization (NMF) is a technique that involves a clear interpretation in terms of Q-matrix, also referred to as transfer model [41]. There are many NMF algorithm, and they can give different solutions. NMF uses an array of positive numbers is the product of two smaller matrices. For example, when a learning process is considered, the matrix may represent the results of students' testing and can be decomposed into two matrices: Q, which represents learning elements and S, representing each student's skills.

The purpose of Process mining is the extraction of knowledge about the process in the learning systems from event logs for the full representation of the entire process, its analysis and improvement. In the EDM Process mining can be used to present the students' behavior according to the records in the event log. Data about each event contain the time stamp and the data about learning process. This may be information about students' knowledge assessment [42], information on participation in forums and chats, about lectures and other educational materials viewing, information about passing tests [43], data describing the collaborative learning processes (the log files of chat and editor) [44], information about events related to the metacognitive prompts [45]. It is important to define the concept of the event (it could be a mouse click) and the concept of the sequence of events. Depending on the behavior of students, they can be combined into different groups. For visualization of individual events Dotted Chart diagrams are often used. Further a construction of process models and conformance checking take place. Building of the learning process model is complicated by the existence of loops and parallel tasks, the presence of "noise", the mutual influence of some tasks to others. The process model is usually presented in the form of a BPMN model or as a Petri net. To construct and test learning process models the general and special Process mining algorithms are used, such as alpha-algorithms, probabilistic, heuristic and genetic algorithms, as well as the Data mining methods and algorithms.

Unfortunately, in the Russian scientific journals, in spite of the considerable amount of work in the field of data mining, there are still little scientific papers related to the study of the application of Data mining and Process mining technology in the learning process. Among them there are the use of methods of Educational Data Mining and Learning Analytics in the educational qualifications [46], the study of the factors of adaptation of students to training conditions with the help of the analysis of variance method [47], an overview of the tasks and methods of Data mining in the field of education and the use of classification algorithms for data analysis of training systems [48], the use of artificial neural networks in the modeling of educational process in high school [49], the study of the structure of high school students values by means of cluster analysis [50].

### *3.3. Scientific conferences and journals in the field of Educational Data Mining*

EDM became an independent research area in recent years. It includes research on the training of intellectual systems - Intelligent tutoring systems (ITS), Artificial intelligence in education (AIED), User modeling (UM), Technology-enhanced learning (TEL), as well as Adaptive and intelligent educational hypermedia (AIEH).

The first conference EDM2008 is held in Montreal, Canada; EDM2009 in Cordoba, Spain; EDM2010 in Pittsburgh, USA; EDM2011 in Eindhoven, the Netherlands; EDM2012 in Chania, Greece, EDM2013 in Memphis, USA, EDM2014 in London, UK, EDM2015 in Madrid, Spain, and EDM2016 in Raleigh, USA. Table 2 summarizes some of the other conferences that correspond to the field of EDM:

**Table 2.** Scientific conferences that correspond to the category EDM

Title	Short title	Type	Starting year
International Conference on Educational Data Mining	EDM	annual	2008
International Conference on Learning Analytics and Knowledge	LAK	annual	2011
International Conference on Artificial Intelligence in Education	AIED	every two years	1983
International Conference on Intelligent Tutoring Systems	ITS	every two years	1988
International Conference on User Modeling, Adaptation, and Personalization	UMAP	annual	2009

Table 3 provides examples of journals corresponding to the field of EDM:

**Table 3.** Examples of journals corresponding to the field of EDM

Title	Short title	Publisher
Journal of Educational Data Mining	JEDM	EDM Society
International Journal of Artificial Intelligence in Education	IJAIED	AIED Society
Journal of the Learning Sciences	J Learn Sci	Taylor&Francis
Computer and Education	CAE	Elsevier
IEEE Transactions on Learning Technologies	TLT	IEEE
IEEE Transactions on Knowledge and Data Engineering	TKDE	IEEE
ACM Special Interest Group on Knowledge Discovery and Data Mining, Explorations	SIGKDD Explorations	ACM
User Modeling and User-Adapted Interaction	UMUAI	Springer
Internet and Higher Education	INTHIG	Elsevier
Journal of Educational and Behavioral Statistics	JEBS	SAGE Publications

Most accurately the theme of the domain is presented in Journal of Educational Data Mining (<http://www.educationaldatamining.org/JEDM/>), published since 2009. Journal of Educational Data Mining is available as an online journal with free access.

#### 4. Conclusion

The paper discusses the basic principles of research in EDM domain, some examples of tasks that can be solved by the use of data mining and Process mining in the area of traditional and e-learning are given, the possibilities and limitations of different methods are described, an overview of the major scientific conferences and journals devoted to the application of Data mining and Process mining techniques in education is presented.

EDM allows investigation on the content of learning materials in e-learning systems and the processes performed in it to be carried out.

The use of Information and Communication Technologies in education generates a large amount of data that contains comprehensive information for students, the processes through which they pass in the course of education. The data derived and used by stakeholders (teachers, instructors, etc.) to understand the learning habits of students, the factors affecting their performance and skills they acquire can be examined. To answer these questions, the research interest in the use of Data mining in education increases. EDM is a discipline aimed at developing specific methods to study educational databases generated by any type of information system supporting training or education (schools, colleges, universities, or vocational training institutions offering traditional and/ or modern methods teaching and informal learning). EDM brings together researchers and practitioners from computer science, education, psychology, psychometrics, and statistics.

The basic idea of Process mining is detecting, monitoring and improvement of real processes by extracting knowledge from event logs automatically recorded by information systems. This approach can be applied to the problems of education. The main goals in this direction are:

- The extraction of process-related knowledge from large education event logs, such as: process models following key performance indicators or a set of curriculum pattern templates.
- The analysis of educational processes and their conformance with established curriculum constraints, educators' hypothesis and prerequisites.
- The enhancement of educational process models with performance indicators: execution time, bottlenecks, decision point, etc.
- The personalization of educational processes via the recommendation of the best course units or learning paths to students (depending on their profiles, their preferences or their target skills) and the on-line detection of prerequisites' violations.

It can be concluded that the use of complementary methods of Data mining and Process mining in e-Learning systems can improve the quality of teaching, increase its availability and effectiveness.

## Acknowledgements

This work is supported by the Bulgarian National Scientific Research Fund under the contract ДФНИ - И02/13.

## References

- [1] Gartner Inc. IT Glossary Web Site [Electronic resource]. — Access mode: <http://www.gartner.com/it-glossary/data-mining> (21.01.2017)
- [2] SAS Institute Inc. Web Site [Electronic resource]. — Access mode: [http://www.sas.com/en\\_us/insights/analytics/data-mining.html](http://www.sas.com/en_us/insights/analytics/data-mining.html) (21.01.2017)
- [3] SIGKDD Web Site [Electronic resource]. — Access mode: <http://www.kdd.org/curriculum/index.html> (21.01.2017)
- [4] van der Aalst, W.M.P. *Process Mining: Discovery, Conformance and Enhancement of Business Processes* / W.M.P. van der Aalst – Berlin: Springer-Verlag, 2011. - 370p.
- [5] Web Site of IEEE Task Force on Process Mining. *Process Mining Manifesto* [Electronic resource]. — Access mode: [http://www.processmining.org/blogs/pub2012/process\\_mining\\_manifesto](http://www.processmining.org/blogs/pub2012/process_mining_manifesto) (21.01.2017)
- [6] Slater, S. *Tools for Educational Data Mining: A Review* / S. Slater, S. Joksimović, V. Kovanovic, RSJd. Baker, D. Gasevic // *Journal of Educational and Behavioral Statistics* – 2017. - Vol. 42, No. 1, - P. 85–106.
- [7] *Handbook of Educational Data Mining* / Edited by C. Romero, S. Ventura, M. Pechenizkiy and RSJd. Baker. – Chapman & Hall/CRC Press, 2011. – 526 p.
- [8] Romero, C. *Data mining in education* / C. Romero, S. Ventura // *The Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* – 2013. - V. 3 - P. 12–27
- [9] Romero, C. E. *Data mining in course management systems: Moodle case study and tutorial* / C. Romero, S. Ventura, E. Salcines // *Comput Edu* - 2008. – V.51 – P. 368–384.
- [10] Mostow, J. *Some useful tactics to modify, map and mine data from intelligent tutors* / J. Mostow, J. Beck // *J Nat Lang Eng* - 2006. - V. 12 - P. 195–208.
- [11] Merceron, A. *Mining student data captured from a web-based tutoring tool: initial exploration and results* / A. Merceron, K. Yacef // *J Interact Learn Res* – 2004. - V. 15 – P. 319–346.
- [12] Romero, C. *Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data* / Romero C, A. Zafrá, J.M. Luna, S. Ventura // *Expert Syst J Knowl Eng*. - 2013. - V. 30(2) – P.162-172.
- [13] Tang, T. *Recommender systems for and in social and online learning environments* / T. Tang, BK. Daniel, C. Romero // *Expert Syst J Knowl Eng*.- 2015. – V.32(2) - P.261-263.
- [14] Romero, C. *Preface to the special issue on data mining for personalised educational systems* / C. Romero, S. Ventura // *User Model User-Adapted Interact* – 2011. – V. 21 – P. 1–3.
- [15] Bannert, M. *Process mining techniques for analyzing patterns and strategies in students' self-regulated learning*. / M. Bannert, P. Reimann, C. Sonnenberg // *Metacognition and Learning*. - 2014. – V. 9(2), P. 161–185.
- [16] Siemens, G. *Learning analytics and educational data mining: towards communication and collaboration* / G. Siemens, RSJd. Baker // *Proceedings of the 2<sup>nd</sup> International Conference on Learning Analytics and Knowledge*. - Vancouver, British Columbia, Canada; 2012. – P. 1–3.
- [17] Kotsiantis, S. *A combinatorial incremental ensemble of classifiers as a technique for predicting student's performance in distance education* / S. Kotsiantis, K. Patriarcheas, MN. Xenos // *Knowl-Based Syst* - 2010. – V. 23 – P. 529–535.
- [18] Frias-Martinez, E. *Survey of datamining approaches to user modeling for adaptive hypermedia* / E. Frias-Martinez, S. Chen, X. Liu // *IEEE Trans Syst Man Cybern C* – 2006. – V. 36, 6 – P. 734–749.
- [19] Pavlik, P. *Learning factors transfer analysis: using learning curve analysis to automatically generate domain models* / P. Pavlik, H. Cen, K. Koedinger // *Int Conf EduDataMin* – 2009. – P. 121–130.
- [20] Bouchet, F. *Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning* / F. Bouchet, J. M. Harley, G. J. Trevors, R. Azevedo // *Journal of Educational Data Mining* – 2013. V. 5(1) – P. 104–146.
- [21] Ayers, E. *A comparison of student skill knowledge estimates* / E. Ayers, R. Nugent, N. Dean // *International Conference On Educational Data Mining*. Cordoba, Spain – 2009. – P. 1–10.
- [22] Cairns, A.H. *Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining* / A.H. Cairns, B. Gueni, M. Fhima, A. Cairns, S. David, N. Khelifa // *The Fourth International Conference on Advances in Information Mining and Management* – 2014. – P. 53-58.

- [23] Garcia, E. Collaborative data mining tool for education / Garcia E, Romero C, Ventura S, Castro C. // International Conference on Educational Data Mining. - Cordoba, Spain, 2009. - P. 299–306.
- [24] Hsia T., Shie A, Chen L. Course planning of extension education to meet market demand by using data mining techniques—an example of Chinkuo Technology University in Taiwan / Hsia T., Shie A, Chen L. // Expert Syst Appl J – 2008. - V. 34 - P. 596– 602.
- [25] Wauters, K. Acquiring item difficulty estimates: a collaborative effort of data and judgment / K. Wauters, P. Desmet, W. Noortgate // International Conference on Educational Data Mining. - Eindhoven, The Netherlands, 2011. – P. 121–128.
- [26] Baker, R. Educational data mining and learning analytics / R. Baker, G. Siemens // Cambridge Handbook of the Learning Sciences: 2nd Edition – 2014. - P. 253-274.
- [27] Romero, C. Web usage mining for predicting marks of students that use Moodle courses / C. Romero, P. Espejo, A. Zafra, J. Romero, S. Ventura // Comput Appl Eng Edu J. - 2013. - V.21 – P.135–146.
- [28] Baker, RSJd. Automatically detecting a student's preparation for future learning: help use is key / RSJd Baker, SM Gowda, AT Corbett // In: Fourth International Conference on Educational Data Mining. - Eindhoven, The Netherlands, 2011. - P. 179–188.
- [29] Bogarín, A. Clustering for improving educational process mining / A. Bogarín, C. Romero, R. Cerezo, M. Sánchez-Santillán // Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. / ACM - New York, NY, USA, 2014 - P. 11-15.
- [30] Vellido, A. Clustering Educational Data / A. Vellido, F. Castro, A. Nebot // Handbook of Educational Data Mining - Boca Raton, FL: Chapman and Hall/CRC Press, 2011. – P. 75– 92.
- [31] Dutt, A. Clustering Algorithms Applied in Educational Data Mining / A. Dutt, S. Aghabozrgi, M.A.B. Ismail, H. Mahrooian // International Journal of Information and Electronics Engineering - 2015. - Vol. 5, No. 2 - P. 112-116.
- [32] Ueno, M. Online outlier detection system for learning time data in e-learning and its evaluation / M. Ueno // International Conference on Computers and Advanced Technology in Education. - Beijing, China, 2004. – P. 248– 253.
- [33] Merceron, A. Measuring correlation of strong symmetric association rules in educational data. / A. Merceron, K. Yacef // Handbook of Educational Data Mining - Boca Raton, FL: CRC Press, 2011. - P. 245–256.
- [34] Rabbany, R. Analyzing participation of students in online courses using social network analysis techniques / R. Rabbany, M. Takaffoli, O. Zaïane // International Conference on Educational Data Mining. - Eindhoven, The Netherlands, 2011. – P. 21–30.
- [35] Tane, J. Semantic resource management for the web: an e-learning application. / J. Tane, C. Schmitz, G. Stumme // International Conference of the WWW. - New York, 2004. - P. 1–10.
- [36] Baker, RSJd. Data mining for education / RSJd. Baker // International Encyclopedia of Education. 3rd ed. Vol. 7. - Oxford, UK: Elsevier, 2010. - P. 112–118.
- [37] Mazza, R. GISMO: a graphical interactive student monitoring tool for course management systems / R. Mazza, C. Milani // International Conference on Technology Enhanced Learning - Milan, Italy, 2004. P. 1–8.
- [38] Baker, RSJd. The state of educational data mining in 2009: a review and future visions / RSJd. Baker, K. Yacef // J Edu Data Min – 2009. P. 3–17.
- [39] Bienkowski, M. Enhancing teaching and learning through educational data mining and learning analytics: an issue brief / M. Bienkowski, M. Feng, B. Means - Washington, D.C.: Office of Educational Technology, U.S. Department of Education, 2012. – P.1–57.
- [40] Corbett, A. Knowledge tracing: modeling the acquisition of procedural knowledge / A. Corbett, J. Anderson // User Model User-Adapted Interact – 1995. – V. 4 - P. 253–278.
- [41] Desmarais, MC. Mapping question items to skills with non-negative matrix factorization. / MC. Desmarais // ACM SIGKDD Explor – 2011. - V. 13 - P. 30–36
- [42] Trčka, N. Process mining from educational data / N. Trčka, M. Pechenizkiy, W. van der Aalst // Handbook of Educational Data Mining. - Boca Raton, FL: CRC Press; 2011. – P. 123– 142.
- [43] Mukala, P. Learning Analytics on Coursera Event Data: A Process Mining Approach / P. Mukala, J. Buijs, M. Leemans, W. van der Aalst // 5th International Symposium on Data-driven Process Discovery and Analysis - Vienna, Austria, 2015. - P. 18-32.
- [44] Schoor, C. Exploring regulatory processes during a computer-supported collaborative learning task using process mining / C. Schoor, M. Bannert // Computers in Human Behavior - 2012. V. 28 - P. 1321–1331.
- [45] Sonnenberg, C. Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques / C. Sonnenberg, M. Bannert // Journal of Learning Analytics - 2015. – V. 2(1) – P. 72–100.
- [46] Veryaev, A.A. Educational Data Mining i Learning Analytics - napravleniya razvitiya obrazovatel'noj kvalilogii / A.A. Veryaev, G.V. Tatarnikova // Prepodavatel' XXI vek. - 2016. - V. 2 P. 150-160. (in Russian)
- [47] Shumetov, V.G. Issledovanie faktorov adaptatsii studenchestva 2000-h godov k obucheniyu v vuze metodami data Mining / V.G. Shumetov, O.V. Lyaskovskaya // Srednerusskij vestnik obshchestvennyh nauk. - 2015. – V.6 – P. 49-56. (in Russian)
- [48] Gorlushkina, N.N. Zadachi i metody intellektual'nogo analiza obrazovatel'nyh dannyh dlya podderzhki prinyatiya reshenij / N.N. Gorlushkina, I.Y. Kocyuba, M.V. Hlopotov // Obrazovatel'nye tekhnologii i obshchestvo - 2015. – V. 1 - P. 472-482. (in Russian)
- [49] Petrova, M.V. Issledovanie vozmozhnostej metodov intellektual'nogo analiza dannyh pri modelirovanii obrazovatel'nogo processa v vuze / M.V. Petrova, D.A. Anufrieva // Vestnik Chuvashskogo Universiteta – 2013. -V. 3 - P. 280-285. (in Russian)
- [50] Avadehni, YU.I. Issledovanie struktury cennostej studentov vuzov s primeneniem tekhnologij data mining / YU.I. Avadehni, O.M. Kulikova, V.A. Radionova // Sovremennye problemy nauki i obrazovaniya. - 2013. – V.6. – P.841 (in Russian)