

REVIEWING PROCESS MINING APPLICATIONS AND TECHNIQUES IN EDUCATION

Athanasios Sypsas and Dimitris Kalles

School of Science and Technology, Hellenic Open University, Patras, Greece

ABSTRACT

Process Mining (PM) emerged from business process management but has recently been applied to educational data and has been found to facilitate the understanding of the educational process. Educational Process Mining (EPM) bridges the gap between process analysis and data analysis, based on the techniques of model discovery, conformance checking and extension of existing process models. We present a systematic review of the recent and current status of research in the EPM domain, focusing on application domains, techniques, tools and models, to highlight the use of EPM in comprehending and improving educational processes.

KEYWORDS

Process Mining, Educational Process Mining, educational applications, process model

1. INTRODUCTION

The widespread use of information and communication systems in educational institutions and training centers supports nearly all educational processes and has led to a vast amount of data collection. Through extensive logging, these systems capture, apart from data, information concerning the supported processes [1]. Data provide insightful information about learners' assessment, learning resource usage, interaction logs, the processes they went through and their involvement in the educational process in general [2], [3]. Additionally, the application of e-learning and blended learning methods, provide helpful information about educational activities at different levels of granularity [4]. Such collected data can be used by educators, learners, institutions, or other educational stakeholders in order to analyze, evaluate and reveal different aspects of the educational process.

Data mining is mainly used to extract useful information from gathered data in order to transform it to a valuable pattern or structure for additional use [5]. Thus, the interest in using data mining techniques in education is constantly increasing aiming to enhance the efficiency and quality of teaching and learning. However, collected educational data are eventually required to be understood and then transformed to actionable information, to achieve impact of some significance [6]. The application of data mining techniques in educational data comes under the term *educational data mining* (EDM). The main goal of this research area is to render most, if not all, of such gathered data useful and valuable to the educational community, with a longer term goal of offering a more personalized, adaptive and interactive educational environment for learners [7].

A related area of research using mainly data mining methods is Learning Analytics (LA) [8]. LA has the potential to detect and clarify characteristics of learning strategies by analyzing captured data [9]. LA combines methods and techniques from other fields such as EDM, process mining

and technology enhanced learning. The generalized use of EDM and LA has led educational institutions to use learning/ content management systems (LMS/ CMS) in order to collect data related to improve the educational process to learners and learning processes [10]. Data mining techniques, such as association analysis and classification have been successfully used with various types of educational data i.e. LMS logs, assessment data, etc. [11]. Although data mining can be used for getting a better understanding of educational processes and providing alternative solutions to the interaction issues during educational process, most of the traditional data mining techniques do not focus on the process as a whole, but on the result [4], [12]. When the focus of the analysis is on the educational process, conventional EDM techniques do not normally lead to meaningful results.

On a related front, process mining (PM) aims to discover, monitor, and improve processes by extracting knowledge from event logs produced by information systems [13], [14]. Thus, PM contributes to obtaining a meaningful process model from event logs helping identify possible problematic issues concerning the investigated process [13]. Even though process mining has emerged from the business community [15], it can also be applied on the EDM context [12]. So, a new research field has emerged, with a focus on discovering and analyzing educational processes, and it has been called Educational Process Mining (EPM) [5]. EPM focuses on the events produced when a process is executed and which are related to completed activities. When PM is applied in an EDM context, it contributes to knowledge and valuable information extraction from educational event log files. Consequently, the research interest on EPM is continuously increasing [2].

In general, a process in an educational context describes the necessary steps which will lead to a specific learning outcome. For example, in virtual reality learning environments process models can be used to abstractly describe the equipment of a physical environment and the experimental processes carried out therein via simulation [16]. Therefore, EPM techniques, when applied on these processes, facilitate the comprehension of experiments and learning flows, since the logged learners' actions are compared with the standardized procedures.

Even though PM is a relatively new technology, the potential of its application in the educational domain is evident in recent research [2]. EPM, apart from data processing and analysis, facilitates decision support and explores possible problematic situations during the educational process. These situations concern, among others, time or performance bottlenecks and non-efficient decision rules that affect the process model. The model extension as a EPM technique based on decision mining, performance analysis and user profiling contributes to detecting and effectively dealing with such situations [3]. The three main types of process mining when applied an educational context are [4], [17]:

1. Process model discovery. The process model is constructed from event logs.
2. Conformance checking. The model is compared to the event log and deviations are detected.
3. Process model extension. The process model is extended or improved using information from the event logs.

All three aforementioned approaches cannot be implemented using traditional EDM techniques. So, new tools and methods were developed in order to extract process-related knowledge from the produced event logs. The application of process mining in the educational field [11] is illustrated in Figure 1 and we briefly review its life cycle below.

The educational information system (i.e., LMS) produces and stores event logs (traces from executed processes). Such events could be activity logs in virtual educational environments (virtual laboratories), learners' registration procedures, access to educational material logs,

Massive Open Online Courses (MOCs), Intelligent Tutoring Systems (ITSs), etc. Using process mining techniques event logs are associated with the process models leading to the extraction of the appropriate information and knowledge concerning process checking, possible execution problems discovering and errors prediction.

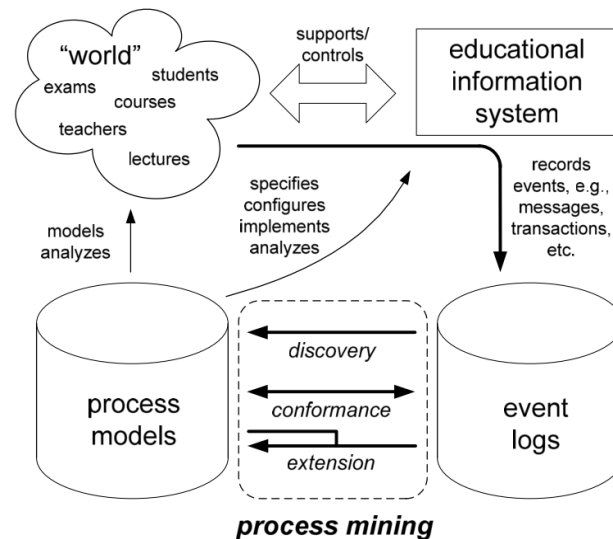


Figure 1. Process mining application in the educational field [23]

However, successful knowledge extraction depends on the quality of the stored event logs, since the huge volume and the heterogeneity of possible traces poses challenges for process mining techniques and algorithms [3], [18], [19]. Using EPM the underlying educational processes can be better understood leading to recommendations and advice to learners, feedback generation for learners, educators and researchers, detecting specific learning disabilities, and improving management of learning objects [20]. The approaches in the present review concern either some or all process mining approaches (Process model discovery, Conformance checking, Process model extension), depending on the application focus and scope.

In this survey, we focus mainly on exploring the various applications of process mining techniques in the education field and, especially, in virtual environments. We aim to provide a comprehensive overview distilling the state-of-the-art from scientific journals and conference proceedings so as to assist researchers assessing recent progress in educational process mining. We have, thus, structured the rest of this paper as follows: In Section 2 the previous work in EPM is summarized, while in Section 3 an overview of research methodology along with research questions is presented. Section 4 outlines the findings with respect to the research questions defined in Section 3 and the final Section 5 provides the conclusion.

2. PREVIOUS WORKS ON EPM

Research on educational process mining is not a new topic. In fact, several reviews have been published concerning process mining in general [13], [17], [21]. However, these reviews do not focus on EPM. Only two other reviews were found that discuss EPM more in depth [2], [22]. Bogarín et al. [2] published a survey on the subject introducing the EPM and elaborating on some of the potential of the specific research area. They presented the components of an EPM framework, the data, tools, techniques and models used in EPM. The papers included were mainly categorized by educational application domains. The major drawback was that it was not

a systematic literature review and the included papers are published by year 2017. Ghazal et al. [22] conducted a literature review collecting 37 papers, containing case studies where EPM was applied. The authors searched for tools and techniques used in EPM and analyzed the selected case studies and their results. Furthermore, they have discovered and highlighted the potential benefits of EPM. However, results concern case studies by the year 2016.

Our work is distinct from these literature reviews, since the results are up-to-date and includes application domains that were not covered before. From our research was concluded that papers published after 2017 are almost the double compared to the papers before 2017. Finally, this study considers research on all aspects of EPM application, including not only educational case studies, but also administration and curriculum issues.

3. RESEARCH APPROACH

This section covers the research approach used in order to obtain as much relevant information as possible concerning the EPM applications. This study employed literature review to extract the various applications in the research field of EPM.

3.1. Research Questions

The key part of a systematic review is the definition of research questions [23]. In the present review paper, the defined research questions contribute in building a knowledge base about the current state of concepts.

RQ1: Where is Educational Process Mining applied?

RQ2: What are the techniques and/or algorithms used in Educational Process Mining research?

RQ3: How is EPM applied on Virtual educational environments?

To answer the aforementioned research questions a search was conducted.

3.2. Conduct Search

In order to execute a literature review, a broad search plan is constructed, so as to ensure the completeness of the search of resources under consideration. A precise search query is needed for scientific databases search, taking into account the related terms extracted from research questions, as can be seen in Table 1, below.

Table 1. Search terms

Key term	Broader terms	Priority
Process mining	Educational process mining, data mining	1
Education	-	2
Techniques	Algorithms	3
Domain	Virtual environment, virtual laboratory, case study	4

Based on the search terms of Table 1, various search strings were tested. After numerous iterations the search string that resulted in the highest number of papers is the following: (“Process mining” AND (“education” OR “virtual laboratory” OR “virtual environment”) AND (“case study” OR “domain” OR “techniques”)). The main source of information and data was the peer-reviewed literature. Searches for peer-reviewed journal articles and dissertations were conducted using the following databases: ACM Digital Library, SCOPUS, Spinger Link and

ScienceDirect. Moreover, sources outside those databases were searched, particularly using the references of relevant articles. Google Scholar was used, but the overlapped articles with other data sources were excluded. The search was performed in June 2021. All the web searches were executed without cookies or login to guarantee any impact avoidance due to previous searches. After the initial search, filtering options offered by digital databases were used in order to apply inclusion (IC) and exclusion (EC) criteria. The inclusion criteria were:

- Peer-reviewed articles published in journals, International Conferences/Workshop Proceedings;
- Full results presentation;
- Frequency of citations;
- Thorough methodology presentation;
- English language researches.

Subsequently, the exclusion criteria were applied on the results. These criteria were:

- Papers not written in English;
- Duplicate papers;
- Editorials or interviews;
- Very short papers, one or two pages long;
- Missing authors papers.

After applying IC and EC 832 articles were remained for review. Publications that obviously were out of the interests of the current study were removed. Then the evaluation focused on abstracts: they were manually filtered by two researchers as relevant to the research criteria and worthy of further study. For the 112 publications that were relevant to our literature search, the full articles were accessed. In the case of any uncertainty concerning a study's relevance based on its title or abstract, we chose to include the study in our full-text evaluation. At the final stage, the full texts of these 112 articles were examined for final worthiness. The resulting literature revealed 95 peer reviewed journal articles and documents. Each document was listed to categorize the information for further analysis, documents were grouped, summarized and critically analyzed.

4. FINDINGS

This section presents findings with respect to the RQs defined in section 3.1.

4.1. Educational Process Mining Application Domains

A wide range of EPM applications in various educational domains is observed. In this subsection the main body of the reviewed literature is presented, giving special significance to the EPM applications and their results.

4.1.1. LMS, LAD and Online courses

Learning Management System (LMS) and Learning Analytics Dashboard (LAD) are online learning environments that offer various learning opportunities and data concerning learning activities. There is a constantly increasing research about the application of EPM in this domain, since these learning environments are used in many educational settings.

In [24] PM techniques were applied to analyse LMS logs from Moodle in order to examine the alignment of teacher's plan and students' use of LMS resources. The results revealed that students' behavior did not fit the plan teachers had in mind. In a recent study [25] institutional Moodle was used as input for PM application and an ad-hoc tool was developed. LMS was also used in [26] to detect effective learning strategies. Using PM techniques researchers concluded that their method could activate learners' awareness about effective learning strategies and motivate them to use them productively. Combining clustering and sequential pattern mining, Martínez-Carrascal et al. in [27], applied generalized sequential pattern (GSP) algorithm on the log data. This technique aimed to detect differences between two groups of students (passing and failing). In an earlier research [28] Moodle was used in order to assess the LMS usage behavior. Authors concluded that process model can be used to tailor the Moodle LMS to the specific needs of the individual courses. The same technique was used in [29]. In [30] dashboards of 35 school teachers were retrieved. PM techniques were used and outcomes showed that teachers paid attention on the group alerts indicating support need. Etinger [31] applied PM techniques in order to discover LMS Course Usage Patterns. Using a new PM algorithm, Inductive Miner (IM), Bogarín et al. [32] tried to discover learning processes under LMS Moodle. As a result, learning model discovery with IM could be a resource for preventing learning failure in LMSs. In a similar environment Ahmad Uzir et al. [33] proposed the use of LMS records to discover time management strategies under self-regulated learning (SRL) approach. Real et al. [34] used PM to verify student learning paths in an Introductory Programming Course. The event logs were collected in the Moodle and the different learning paths were revealed. Along similar lines, Romero et al. [5] proposed the adoption of clustering to improve EPM. They used Moodle log files to optimize both the fitness and comprehensibility of the model providing useful information to educators about student learning and the failing course risk. Bogarín et al. [35] based on data from LMS logs managed to obtain more specific and accurate models of students' behaviour. The online modeling platform BJUTModeling was used in studies [36] and [37] as a tool for student learning evaluation effect. Based on PM, authors achieved student modeling patterns mining and their behaviors interpretation. In an early study [38], PM contributed in exploring regulatory processes under Moodle. Similarly, in [39] Moodle log files were used and PM was proved to be helpful for tutors to effectively cope with the study process of their students. In a study [40] with 1.500 participants in distance learning domain, event logs generated by LMS served as input for PM analysis. Results suggested that applying either classical process mining or modern data mining techniques could offer valuable benefits for this domain.

Cerezo et al. [12] explored students' SRL processes, during an online course by using PM Techniques. They managed to discover models with good fitness values, meaning that students' interactions on the Moodle platform were correctly reproduced. Based on Moodle log files Pechenizkiy et al. [11] applied process mining techniques to online assessment data, illustrated some of the potential PM techniques offer. Likewise, in [41] PM was used to understand students' approaches to learning programming. They used a combination of methods (data from more than one source, process and sequence mining) to gain an in-depth understanding of the programming learning. Shabaninejad, et al. [42] used LAD as source of rich and complex data about students and their learning activities. The learning processes of students were visualized and compared giving insights to educators that could be used to adjust educational process. In a recent research [43] a method was proposed to discover student's activities, navigation paths, and behavior in LMS Moodle. PM application offered the ability to distinguish different variants (patterns) of student activity, helping educators to detect problematic issues during learning process. The same LMS was selected for another study [44], in order to explore student behavior. Specifically, students' interactions while engaging in specific quiz-based learning activities were analyzed contributing in research about students' quiz-taking strategies in LMS. In an online course on system dynamics Reimann et al. [45] used PM to discover models from interaction data in LMS.

4.1.2. MOOCs

Massive Open Online Courses offer free, at most cases, learning opportunities to the online community. Due to the big number of participants, the produced log files serve as input for PM techniques leading to valuable research on the area.

The research [46] involved log data from 16.224 learners, participated in a MOOC. Applying PM and clustering techniques researchers examined the relation between learning behavior and learning progress, offering guidance for educators. In another study [9] PM was used on the trace data originating from learners' interactions with the learning activities in a MOOC to explore the association of learning strategies and the learners' personality traits. Bernal et al. [47] in a MOOC under Coursera platform used PM on activity data from the MOOC log files to extend the knowledge about students' behavior in online environments. In order to study dropout prediction in MOOCs process and sequence mining techniques were applied by Deeva et al. [48] on event data gathered from the edX platform. EdX platform was also used to study educational process intelligence (EPI) [49] providing a contextual-based and informed learning process analysis and decision making through the PM approach. A FutureLearn MOOC was utilized as data resource, in order to investigate variation in learning processes [50]. PM discovery method showed that insights from learners' behavior can inform MOOC designers about adjusting learning activities to different groups of learners based on their goals. Three MOOCs offered resources for exploration, through PM application, of the relationship between theoretical SRL strategies and behavioural patterns on MOOC platforms [51]. Finally, Rizvi et al. [52] worked with logged data from an Open University MOOC offered via FutureLearn. The study compared behaviors concerning participation of both groups with expected learning behavior for all types of learning activities.

4.1.3. E-learning Environments, Micro-Learning Environments, Intelligent Learning Environments and Intelligent Tutoring Environments

Various e-learning tools have been designed and developed for learning activities. Beheshitha, et al. [53] proposed the application of nStudy, a learning tool offered as a Firefox add-on, in order to capture log data of learners' activities within the environment. PM was used to examine the relationship between students' self-reported aptitudes and how they self-regulate their learning in action. A computer-based context, nBrowser, was applied in a recent study [54] to examine how teachers enact SRL as they move from the beginning to later stages, by applying PM on logged data. Okoye et al. [55] applied PM to discover sets of recurrent behaviors that can be found within a learning process. Using activity logs within the learning process, the authors determined suitable learning patterns by means of semantic reasoning which can then be used to address the problem of adapting learning to the captured user profiles.

Students' work in Open-ended Learning Environments (OELEs) derives a holistic view of students' learning behaviors. In [56] PM was applied on logged actions from Betty's Brain learning environment and discovered that high performing students not only execute more actions, but their actions are also supported and effective. Intelligent Learning Environments (ILEs) offer useful data for learning behaviors. So, Jeong et al., [57] in a rather early study, took advantage of them to analyze, via PM, the differences in transitional state behavior between the high and the low- performing learners. In a hypermedia learning environment, serving as a computer-based learning environment (CBLE), process model discovery and conformance checking were used to investigate both the effect of metacognitive prompts on metacognitive and cognitive learning processes and the SRL process itself [58]. An e-Learning environment (DME) for software operation data offered the log framework XES as input for PM technique of conformance checking [59]. Authors reported that defining and obtaining the right software

operation data is a process and PM contributed in process analysis. Deeva et al. [48] analyzed behavioral data from the JMermaid modeling environment to explore potential behavioral patterns and indicative features correlated with better learning outcomes. The same modeling environment was selected for another study [60]. The interactions with the modeling tool have been logged and PM was used to examine how educators endorse SRL as they move from the beginning to later stages of designing lessons. An intelligent tutoring system offered input data for PM application in a research [61], scoping to characterize learner behaviors at different phases of problem solving. Deeva and Weerdts [62] used a Smart Learning Environment. Various PM techniques were applied leading to student interactions capture in order to improve learning experiences. Poncin et al. [63] made use of PM techniques on Capstone projects educational approach to obtain insights in the development process followed. In a study [64] a cloud-based Educational Software Tool was developed to offer resources for PM analysis aiming at understanding and validating the behavior of team projects in capstone courses. Micro-Learning is a new learning approach based on e-mail or SMS, chat rooms, microblogging in which a primary learning resource is split in a series of micro learning units that are distributed via the Internet. Chen et al. [65] proposed the application of PM to support personalized micro-learning by spotting and extracting the suitable micro-learning units and then organize them as a suitable learning process matching the learners' needs.

A hypermedia learning environment worked as the data collector for a study [66]. Model discovery and conformance checking were combined to analyze the temporal order of spontaneous individual regulation activities and how findings can be applied to identify process patterns in SRL events. Hypermedia learning sessions were used in [67], applying PM to detect long-term effects of metacognitive prompting on SRL in a follow-up learning task without instructional support. Combining the Interactive Digital Narratives (IDN) approach under Prolific, a scientific crowdsourcing platform and PM, Estupiñán and Szilas [68] examined the understanding of user engagement, particularly, spotting when and what happened when user engagement dropped.

4.1.4. Flipped Classroom approaches and blended learning approaches

The combination of e-learning and face-to-face educational approaches offers a wide range of applications resulting in learner-friendly environments. These approaches are utilized to extract valuable information when PM techniques are applied.

In a recent study [69] researchers proposed PM usage in event logs derived from project-based, collaborative undergraduate engineering course to illuminate how regulation phases can facilitate subsequent regulation. In [70] a three-year study was conducted, applying PM techniques on the data collected from class preparation activities of a flipped classroom targeting to reveal useful information from trace data and the way they are associated with academic performance and feedback. Similarly, in another study [71] concerning a course that followed a flipped classroom pedagogy to study SRL, PM was used for a richer comprehension into SRL behaviors. The application of PM at a flipped classroom approach aiming at providing real-time or after-class indicators to students and educators was studied in [72]. At a similar flipped classroom approach Saint, Gašević et al. [73] discussed the PM use for learning strategies detection, concluding that more strategically minded activity, embodying learner SRL, generally proves to be more successful than less disciplined reactive behaviors. Likewise, in [74] the PM application to support SRL under flipped classroom environment was explored. Bolt et al [75] combined the complementary approaches of process cubes and analytic workflows in the field of PM to analyze large, heterogeneous event data of multidimensional nature. Logged data offered resources for an older study [76], where PM application implemented the comparison of the learning processes inside and between courses contributing in better understanding of the study behavior of students.

Heirweg et al. [77] contrary to the very limited existing research in this regard, applied PM on online data collection through thinking aloud method exploring the sequence in which SRL activities are conducted during learning. Under a blended-learning approach PM was used to identify process patterns as part of the tailored feedback on medical procedural training [78]. Dahari et al. [79] extended a PM tool (ProM) framework for educational purposes in order to achieve a rich development environment for executing, creating, and extending mining techniques.

4.1.5. Online Collaborative Environments

Online collaborative environments such as wikis and social networks pose a challenge for PM application, since the mined data and logged events can be easily capture for further analysis.

Virtual collaborative system KP-Lab System is used in an early study [80] where a standalone tool served for PM application, via critical patterns recognition, on the data logged by the system aiming to study learners' activities. Ten years later Chen and Wang [81] studied the use of a Collaborative Web-based Inquiry Science Environment (CWISE) as data collector for PM application in order to study effective learning behaviors. In the same way [82] the PM applied on event logs collected and captured during a distance learning activity in a computer-based collaborative learning environment.

Social networks via group formation offer an online collaborative space for educational processes. Thus, Peeters et al. [83] exploited Facebook as an online collaborative environment for foreign language learners. Through PM the researchers analyzed a range of SRL tactics and task phases followed by learners to manage their academic writing process. Another social network (WhatsApp) served as the collaboration platform in [84] where PM techniques included data pre-processing, discovering process, and conformance checking provided perceptions that can be used to improve student collaboration.

Although in recent years wikis are not used frequently, they can serve as a repository for learning activities. Hence, in an early study [85] PM was applied on activity logs generated by process aware information systems such as Wiki based document management system, to help the educator in giving directed and specific feedback. Caballero-Hernández et al. [86] suggested the use of PM techniques on wiki log data to analyse the behavior of team members in sequential processes like collaborative projects.

4.1.6. Online collaborative environments

Game Based Learning (GBL) is the educational approach where serious games are used for educational purposes. The main purpose of entertainment games is the fun of the player while serious games have educational purposes. Hernández [87] used the PM technique of event sequence analysis to study skill assessment in learning experiences based on serious games. PM application on events in MySQL database of a game-based assessment tool, Shadowspect, revealed how students interacted with the game and spotted the problems that the students were having while playing [27]. Authors in [88] used prototype metrics to process and analyze students' self-reflection in the context of Radix, a multi-player inquiry-based learning game on STEM topics scoping to expand the game itself into a more complete game-based learning system.

Finally, a research [89] explored the possibilities of application of PM techniques in the analysis of interactions among players, and the involved decision taking processes, in non-digital games.

4.1.7. Curriculum and carrier analysis

One of the important steps to ensure that a syllabus aligns with the educational setting's mission as well as the education standards from the government is curriculum assessment. When curriculum model is taken under consideration, students have a guideline to consider the learning process during their study period [90]. Having this in mind, the abovementioned researchers aimed at investigating the actual curriculum model by constructing the model from the students' academic results using PM techniques. Based on the curriculum analysis when sequence matching analysis via conformance checking is performed, another study [91] evaluated the a-priori curriculum design with actual curriculum learning sequence. In one of the earliest studies [92] a new domain-driven EPM framework was proposed to facilitate more focused search for local patterns based on logged data from academic curriculum.

As far as carrier analysis is concerned, Cameranesi et al. [93] utilized the set of traces representing students' careers from ESSE3, a software suite commonly adopted by Italian Universities to manage students' careers and performed PM application that led to a preliminary analysis of students' careers.

4.1.8. Students' enrolment and performance assessment

A long-term research which aimed at studying student's enrolment used data was based on anonymized enrolment and course-grade records for 6 cohorts of students between 2004 and 2009 [94]. PM discovery algorithms, combined with statistical techniques, provided visual representations that allow an easy-to-understand analysis, helping to comprehend the process that led undergraduate students to drop out. Salazar-Fernandez et al. [95] made use of the same educational trajectories of engineering students in each one of the high-failure rate courses as the previous study. PM analysis of trajectories allowed researchers to confirm the importance of prior time and resources investment regarding the students' decision to continue their studies. A data set with relevant information about the performance of 612 students was the source for PM application in [96]. The analysis of students' performance revealed courses which possibly represent bottlenecks in the process of completing a computer science degree.

Online Assessment Test is an evaluation form of a student learning process. Aisa et al. [97] evaluated the online assessment test via PM application so as to extract the frequency of each task ID revealing the strengths and weaknesses of each student. In a similar study, workplace based assessment (WBA) was suggested as a tool to collect feedback on students' practical skills, their performance, and goals setting [98].

4.1.9. Professional training

Many organizations and companies invest capitals on their employees training targeting at their professional development. From our research five studies exploited the PM application for professional training. Ariouat et al. [20] suggested a two-step clustering PM technique on real-world professional training databases from a worldwide consulting company. Their approach was adequate to partition efficiently training event logs in order to achieve a better understanding of the underlying educational processes. In a 6.000 employees study [99], semantic annotations to educational event logs were associated. Thus, more accurate and compact educational processes were extracted and analyzed at different levels of abstraction providing useful information for employees' professional training. Cairns et al. [100] used a process log, issued from the Altran training institute activities. Authors used sequence clustering and then PM techniques which contributed in the development of a distributed platform dedicated to educational process model discovery and conformance checking.

In Medical Training domain, a recent research [78] studied the PM application on surgical procedures. As a result, tailored feedback of surgical procedures using a process perspective was mined. In the same domain the basic concepts to design medical training applications using PM was introduced [101]. The proposed methodology Process-Oriented Medical Education (POME) was applied on a running case of a surgical procedure, which has been used to develop applications in medical training using PM.

4.1.10. Educational Virtual Environments

Educational Virtual Environments are offering an educational approach, where learners interact with educators or with other learners in a virtual environment. Additionally, in some cases, they implement a virtual world where students are conducting experiments which could not be executed in a real life setting i.e., not accessible chemistry or biology laboratories. In such environments, PM has been used to analyze the learning processes taking place there. Sypsas and Kalles [102] proposed the application of conformance checking PM technique on logged events from a virtual biology laboratory (Onlabs). The authors suggested that the research outcomes assist educators to decide whether the specific virtual lab can be used as an educational tool in various settings i.e., universities, secondary education schools, etc. A 3D biotechnology virtual lab combined with an Intelligent Tutoring System (ITS) is used in another study [103]. PM on student logs were used to implement an automatic tutoring adapted to each type of student by applying methods of extraction and analysis of data. Two other studies presented the PM usage on data taken out from logged events inside virtual learning environments [104], [105]. Units of learning (UoLs) from different domains have been used to extract the learning flow structure and to obtain the underlying rules that control the adaptive learning of students. Proposed approaches claimed that facilitated the reuse of UoLs defined in virtual learning environments. Castillo [106] suggested the PM use in the context of a virtual laboratory for a multiagent programming course to detect successful and failed behaviors of students. Another simulation-based environment TELEOS (Technology Enhanced Learning Environment for Orthopedic Surgery) served as a data source in [107]. The extraction of sequential rules common to several sequences was used to mine the distinctive phases of simulated operations of vertebroplasty.

Finally, Fernández-Gallego et al. [108] used an Opensim-based virtual world platform, known as OPENET4VE, to develop applications that automatically extract the learning flow undertaken by students in virtual worlds.

4.1.11. Real World Processes

In [109] a novel PM technique for linguistic summarization of event logs was suggested, which generated linguistic summaries that are concise enough to be used in a practical setting. Execution log of a Dutch municipality process was utilized and results were proved to be practically usable by a human interpreter.

In conclusion, in Figure 2 the number of each application is presented. The most active domains are LMS, LAD and Online courses, E-learning Environments, Micro-Learning Environments, Intelligent Learning Environments and Intelligent Tutoring Environments, MOOCs and Flipped Classroom and blended learning approaches. Some studies (4) can be classified in two domains as they use approaches from more than one.

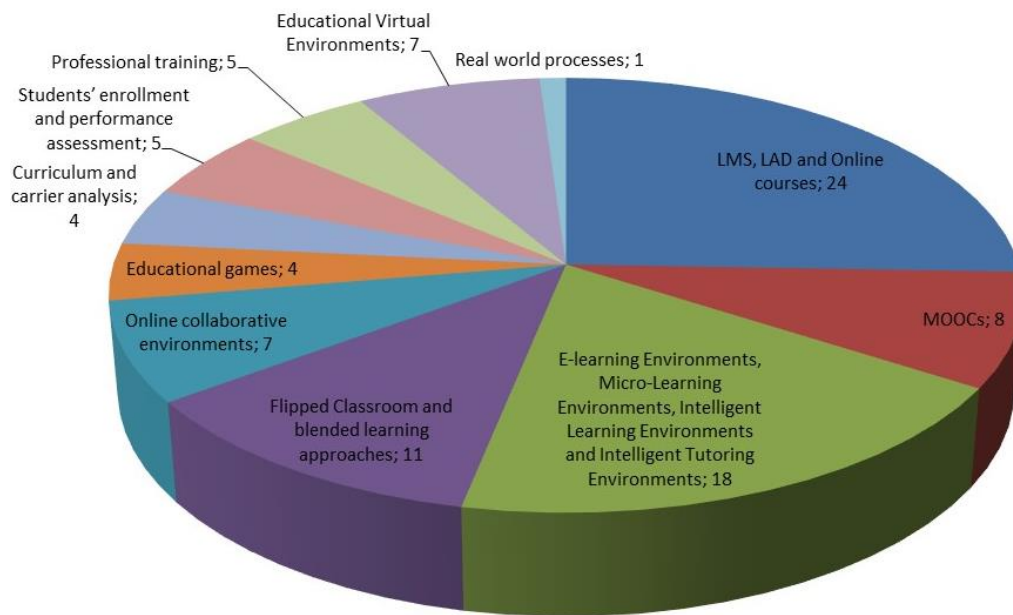


Figure 2. EPM studies per domain

4.2. Techniques and/or Algorithms used in Educational Process Mining Applications

Most studies in the present paper used PM techniques and algorithms that have already been implemented in PM tools such as ProM and Disco. However, some researchers combined techniques and algorithms from Data Mining and Learning Analytics domains with PM in order to achieve better results. Additionally, they developed plug-ins in existing tools or new tools which implement these approaches. In Figure 3 below, the PM tools are presented, showing that the open-source tool ProM and the Disco (Fluxicon) tool dominate in EPM applications.

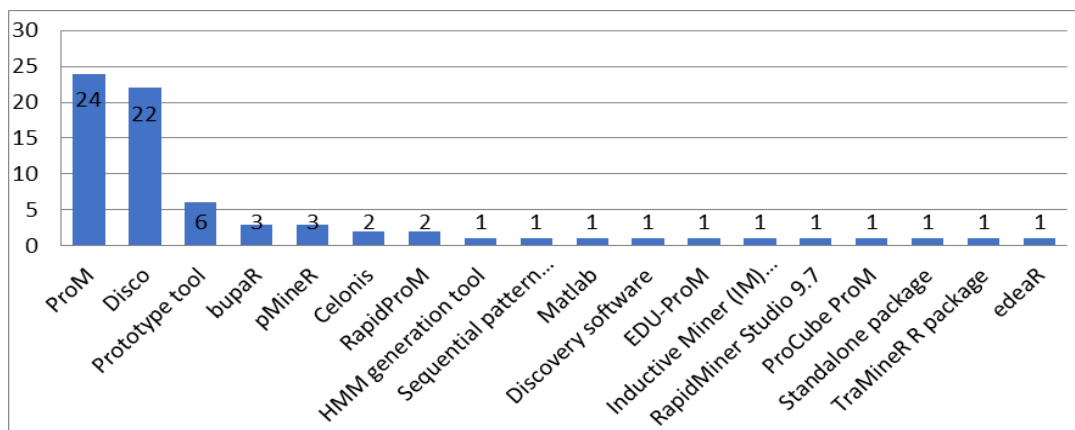


Figure 3. PM tools used in EPM studies

The model discovery and conformance checking techniques are used almost in all studies. In a rather big number of studies more than one algorithm and techniques are used targeting at finding the optimal solution. Figure 4 depicts the algorithms and techniques that are used in the studies included in our research. The most commonly used algorithms are Fuzzy miner (23.62%),

Heuristic miner (18.89%) and Inductive miner (11.02%), while Conformance checking is adopted by 13 researches.

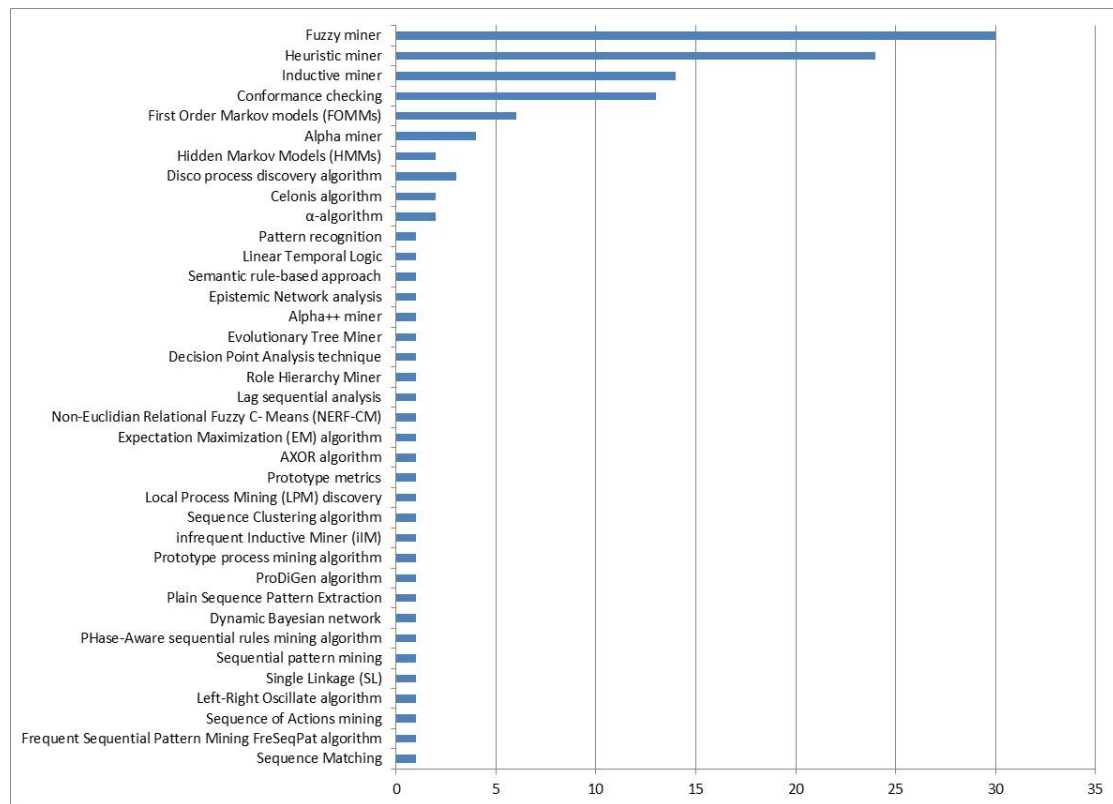


Figure 4. PM algorithms and techniques used in EPM

Model discovery and conformance checking are using model representation. The most used representations from the studies where process map (21,11%), Petri nets (17,78%), process model (10%) and various kinds of Graphs (cyclic, directed, weighted, etc.) (10%).

4.3. EPM on Virtual Educational Environments

Virtual educational environments have been adopted as educational tools during the last decade, since they offer an attractive and interactive learning environment. The most used applications on the domain are the virtual laboratories and the virtual worlds. In these applications process models are used to represent both the virtual learning environment i.e., laboratory equipment and the learners' steps, i.e., experimental steps. The seven studies using EPM in virtual educational environments are presented in section 4.1.10. Four out of seven studies concern virtual laboratories, two virtual laboratories for biology, biotechnology, one for medical training and one for computer science. Two studies are based on Virtual Learning Environments and one on virtual worlds. EPM techniques applied are process discovery and conformance checking based on the log files produced by the virtual environments. Four papers used Petri nets for model representation, while the other three Bayesian network, extended automata and process map. As far as EPM tools are concerned the majority (57, 15%) of researches adopted prototype tools even though there are free tools available (ProM).

5. CONCLUSIONS

Even though PM was firstly applied in more traditional business domain, the lack of a process-centric approach concerning educational data has spearheaded work on exploring the applicability of PM to the education domain, leading to the emergence of EPM [110]. The present research review introduces EPM and provides an overview of its applications in various educational domains offering researchers, educators and institutions a view of the different approaches over the last decade. Additionally, tools, techniques and algorithms are presented to contribute to a deeper comprehension of EPM and its applications. The analysis of the selected studies exemplified the approaches and tools used by researchers, and highlighted the potential benefits of EPM. Most applications concern online courses and e-learning environments, since the used tools (LMS) offer logging capabilities for learners' data and procedures. Furthermore, many papers suggest the use of EPM combined with SRL leading to learners' behavior study. Thus, educators and institutions are able to detect possible problematic situations during educational process, mainly in distance education courses, where SRL is a prerequisite.

EPM, as an evolving research area, faces many challenges for the future. A first one is the improvement of model discovery, since the model is used to abstractly describe all possible learners' actions and processes taking place during the educational procedure. Thus, improved model discovery may capture as many processes as possible for further EPM analysis. Secondly, semantics could be used to provide a more theoretical analysis of the processes taking place in educational procedure. This may lead to the design of more detailed log generation application in the existing information systems as semantics are associated with the logged events. This approach was also mentioned in a study [3] where labels were suggested to be linked with events. Thirdly, the results of EPM application should be formed in order to be understandable from all the involved parts, educators, learners, institutions, stakeholders. This, could serve as an initiative for the development of interactive environments with smart visualization techniques contributing on a more effective and efficient use of EPM [3]. Lastly, EPM tools and EPM datasets could be freely available so that researchers can use them to test their approaches offering a boost on EPM research.

As far as virtual laboratories are concerned, the potential of EPM applications is not fully explored. Irritative challenges among others are the implementation of applications that efficiently log processes from as many experiments from various domains as possible and the detection of learners' actions inside the lab that need improvement compared with learning outcomes that have been set. This could be achieved through the incorporation of a virtual lab assistant or chatbot.

This review aims at providing the educational community and researchers with the up-to-date EPM techniques and applications on educational data. EPM facilitates a better comprehension of the underlying educational process from raw event data contributing in effective learning results. However, special attention should be paid on AI applications, as EU guidelines (https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682) identify AI systems as high-risk when used in Educational or vocational training, as it may determine the access to education and professional course of someone's life (e.g., scoring of exams).

REFERENCES

- [1] A. A. Kalenkova and D. A. Kolesnikov, "Application of a Genetic Algorithm for Finding Edit Distances between Process Models," *Autom. Control Comput. Sci.*, vol. 53, no. 7, pp. 617–627, 2019.
- [2] A. Bogarín, R. Cerezo, and C. Romero, "A survey on educational process mining," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 1, p. e1230, 2018.

- [3] A. H. Cairns, B. Gueni, M. Fhima, A. Cairns, S. David, and N. Khelifa, "Process Mining in the Education Domain," *Int. J. Adv. Intell. Syst.*, vol. 08, no. 1 & 2, pp. 219–232, 2015.
- [4] N. Trčka, M. Pechenizkiy, and W. van der Aalst, "Process mining from educational data," in *Handbook of Educational Data Mining*, Chapman &., C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, Eds. CRC Press, 2010, pp. 123–142.
- [5] C. Romero, R. Cerezo, A. Bogarín, and M. Sánchez-Santillán, "EDUCATIONAL PROCESS MINING: A TUTORIAL AND CASE STUDY USING MOODLE DATA SETS," in *Data Mining and Learning Analytics: Applications in Educational Research*, S. ELATIA, D. IPPERCIEL, and O. R. ZAIANE, Eds. John Wiley & Sons, Inc., 2016, pp. 3–28.
- [6] R. Tsoni, E. C. Stavropoulos, and V. S. Verykios, "Leveraging Learning Analytics with the Power of Words," *Envisioning Rep. Empower. Univ.*, vol. 24, pp. 24–26, 2019.
- [7] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telemat. Informatics*, vol. 37, no. January, pp. 13–49, 2019.
- [8] K. H. Kyritsi, V. Zorkadis, E. C. Stavropoulos, and V. S. Verykios, "The pursuit of patterns in educational data mining as a threat to student privacy," *J. Interact. Media Educ.*, vol. 2019, no. 1, pp. 1–10, 2019.
- [9] W. Matcha *et al.*, "Analytics of learning strategies: The association with the personality traits," *ACM Int. Conf. Proceeding Ser.*, pp. 151–160, 2020.
- [10] M. M. Tair and A. M. El-Halees, "Mining Educational Data to Improve Students' Performance: A Case Study," *Int. J. Inf. Commun. Technol. Res.*, vol. 2, no. 2, pp. 140–146, 2012.
- [11] M. Pechenizkiy, N. Trčka, E. Vasilyeva, W. Van Der Aalst, and P. De Bra, "Process mining online assessment data," in *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, 2009, no. June 2014, pp. 279–288.
- [12] R. Cerezo, A. Bogarín, M. Esteban, and C. Romero, "Process mining for self-regulated learning assessment in e-learning," *J. Comput. High. Educ.*, vol. 32, no. 1, pp. 74–88, 2020.
- [13] C. dos S. Garcia *et al.*, "Process mining techniques and applications – A systematic mapping study," *Expert Syst. Appl.*, vol. 133, pp. 260–295, 2019.
- [14] W. Van Der Aalst, "Process mining," *Commun. ACM*, vol. 55, no. 8, pp. 76–83, 2012.
- [15] W. Van Der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142, 2004.
- [16] J. A. Rossiter, "Low production cost virtual modelling and control laboratories for chemical engineering students," *IFAC-PapersOnLine*, vol. 49, no. 6, pp. 230–235, Jan. 2016.
- [17] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *J. Biomed. Inform.*, vol. 61, pp. 224–236, 2016.
- [18] R. P. J. C. Bose, R. S. Mans, W. M. P. Van Der Aalst, B. F. A. Hompes, H. M. W. Verbeek, and W. M. P. Vanderaalst, "Wanna improve process mining results?," *Proc. 2013 IEEE Symp. Comput. Intell. Data Mining, CIDM 2013 - 2013 IEEE Symp. Ser. Comput. Intell. SSCI 2013*, vol. 237, pp. 32–57, 2015.
- [19] B. F. A. Hompes, H. M. W. Verbeek, and W. M. P. Van deraalst, "Finding suitable activity clusters for decomposed process discovery," in *Lecture Notes in Business Information Processing*, 2015, vol. 237, pp. 32–57.
- [20] H. Ariouat, A. H. Cairns, K. Barkaoui, J. Akoka, and N. Khelifa, "A two-step clustering approach for improving educational process model discovery," *Proc. - 25th IEEE Int. Conf. Enabling Technol. Infrastruct. Collab. Enterp. WETICE 2016*, no. June, pp. 38–43, 2016.
- [21] D. Dakic, D. Stefanovic, T. Lolic, D. Narandzic, and N. Simeunovic, "Event Log Extraction for the Purpose of Process Mining: A Systematic Literature Review," in *15th International Symposium in Management*, 2020, no. October, pp. 299–312.
- [22] M. A. Ghazal, O. Ibrahim, and M. A. Salama, "Educational process mining: A systematic literature review," in *Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017*, 2018, pp. 198–203.
- [23] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.
- [24] P. Alvarez, J. Fabra, S. Hernandez, and J. Ezpeleta, "Alignment of teacher's plan and students' use of LMS resources. Analysis of Moodle logs," *2016 15th Int. Conf. Inf. Technol. Based High. Educ. Training, ITHET 2016*, 2016.

- [25] A. Balderas and J. A. Caballero-Hernández, “Analysis of Learning Records to Detect Student Cheating on Online Exams: Case Study during COVID-19 Pandemic,” in *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'20)*, 2020, pp. 752–757.
- [26] N. A. A. Uzir, D. Gašević, J. Jovanovic, W. Matcha, L. A. Lim, and A. Fudge, “Analytics of time management and learning strategies for effective online learning in blended environments,” *ACM Int. Conf. Proceeding Ser.*, pp. 392–401, 2020.
- [27] M. J. Gomez, J. A. Ruipérez-Valiente, P. A. Martinez, and Y. J. Kim, “Exploring the Affordances of Sequence Mining in Educational Games,” in *TEEM'20: Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2020, pp. 648–654.
- [28] D. Etinger, T. Orehovački, and S. Babić, “Applying process mining techniques to learning management systems for educational process model discovery and analysis,” in *Advances in Intelligent Systems and Computing*, 2018, vol. 722, pp. 420–425.
- [29] N. Kelly *et al.*, “Combining event- and variable- centred approaches to institution- facing learning analytics at the unit of study level,” *Int. J. Inf. Learn. Technol. Artic. Inf.*, vol. 34, no. 1, pp. 63–78, 2017.
- [30] A. Van Leeuwen and N. Rummel, “Comparing teachers’ use of mirroring and advising dashboards,” in *ACM International Conference Proceeding Series, LAK'20*, 2020, pp. 26–34.
- [31] D. Etinger, “Discovering and mapping LMS course usage patterns to learning outcomes,” in *IHSI 2020*, 2020, vol. 1131 AISC, pp. 486–491.
- [32] A. Bogarín, R. Cerezo, and C. Romero, “Discovering learning processes using inductive miner: A case study with learning management systems (LMSs),” *Psicothema*, vol. 30, no. 3, pp. 322–329, 2018.
- [33] N. Ahmad Uzir *et al.*, “Discovering Time Management Strategies in Learning Processes Using Process Mining Techniques,” in *Transforming Learning with Meaningful Technologies. EC-TEL 2019. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11722 LNCS, no. September, pp. 555–569.
- [34] E. M. H. Real, E. Pinheiro Pimentel, L. V. De Oliveira, J. Cristina Braga, and I. Stiubiener, “Educational Process Mining for Verifying Student Learning Paths in an Introductory Programming Course,” in *Proceedings - Frontiers in Education Conference, FIE*, 2020, vol. 2020-Octob.
- [35] A. Bogarín, C. Romero, R. Cerezo, and M. Sánchez-Santillán, “Clustering for improving Educational process mining,” in *ACM International Conference Proceeding Series*, 2014, pp. 11–15.
- [36] Y. Wang, T. Li, C. Geng, and Y. Wang, “Evaluating Student Learning Effect Based on Process Mining,” in *Applied Informatics. ICAI 2019. Communications in Computer and Information Science*, 2019, pp. 59–72.
- [37] Y. Wang, T. Li, C. Geng, and Y. Wang, “Recognizing patterns of student’s modeling behaviour patterns via process mining,” *Smart Learn. Environ.*, vol. 6, no. 1, 2019.
- [38] C. Schoor and M. Bannert, “Exploring regulatory processes during a computer-supported collaborative learning task using process mining,” *Comput. Human Behav.*, vol. 28, no. 4, pp. 1321–1331, 2012.
- [39] K. Slaninová, J. Martinovič, P. Dráždilová, and V. Snašel, “From Moodle Log File to the Students Network,” in *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13*, 2014, vol. 239, pp. 641–650.
- [40] A. R. C. Maita, M. Fantinato, S. M. Peres, L. H. Thom, and P. C. K. Hung, “Mining unstructured processes: An exploratory study on a distance learning domain,” in *Proceedings of the International Joint Conference on Neural Networks*, 2017, pp. 3240–3247.
- [41] S. López-Pernas, M. Saqr, and O. Viberg, “Putting it all together: Combining learning analytics methods and data sources to understand students’ approaches to learning programming,” *Sustain.*, vol. 13, no. 9, 2021.
- [42] S. Shabaninejad, H. Khosravi, S. J. J. Leemans, S. Sadiq, and M. Indulska, “Recommending insightful drill-downs based on learning processes for learning analytics dashboards,” in *International Conference on Artificial Intelligence in Education*, 2020, vol. 12163 LNAI, no. June, pp. 486–499.
- [43] R. Dolak, “Using Process Mining Techniques to Discover Student’s Activities, Navigation Paths, and Behavior in LMS Moodle,” in *Second International Conference, ICITL 2019*, 2019, vol. 11937 LNCS, pp. 129–138.
- [44] L. Juhaňák, J. Zounek, and L. Rohlíková, “Using process mining to analyze students’ quiz-taking behavior patterns in a learning management system,” *Comput. Human Behav.*, vol. 92, pp. 496–506,

- 2019.
- [45] P. Reimann, J. Frerejean, and K. Thompson, "Using process mining to identify models of group decision making in chat data," in *Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1 (CSCL'09)*, 2009, pp. 98–107.
- [46] A. van den Beemt, J. Buys, and W. van der Aalst, "Analysing structured learning behaviour in Massive Open Online Courses (MOOCs): An approach based on process mining and clustering," *Int. Rev. Res. Open Distance Learn.*, vol. 19, no. 5, pp. 38–60, 2018.
- [47] F. Bernal, J. Maldonado-Mahauad, K. Villalba-Condori, M. Zúñiga-Prieto, J. Veintimilla-Reyes, and M. Mejía, "Analyzing Students' Behavior in a MOOC Course: A Process-Oriented Approach," in *HCI 2020: HCI International 2020*, 2020, vol. 12425 LNCS, pp. 307–325.
- [48] G. Deeva, M. Snoeck, and J. De Weerd, "Discovering the impact of students' modeling behavior on their final performance," in *The Practice of Enterprise Modeling. PoEM 2018. Lecture Notes in Business Information Processing*, 2018, vol. 335, pp. 335–350.
- [49] K. Okoye and S. Hosseini, "Educational process intelligence: A process mining approach and model analysis," in *IBICA 2019*, 2021, vol. 1180 AISC, pp. 201–212.
- [50] S. Rizvi, B. Rienties, J. Rogaten, and R. F. Kizilcec, "Investigating variation in learning processes in a FutureLearn MOOC," *J. Comput. High. Educ.*, vol. 32, no. 1, pp. 162–181, 2020.
- [51] J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, "Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses," *Comput. Human Behav.*, vol. 80, pp. 179–196, 2018.
- [52] S. Rizvi, B. Rienties, and J. Rogaten, "Temporal dynamics of MOOC learning trajectories," *ACM Int. Conf. Proceeding Ser.*, no. October, 2018.
- [53] S. S. Beheshitha, D. Gašević, and M. Hatala, "A process mining approach to linking the study of aptitude and event facets of self-regulated learning," *ACM Int. Conf. Proceeding Ser.*, vol. 16-20-Marc, pp. 265–269, 2015.
- [54] L. Huang and S. P. Lajoie, "Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development," *Comput. Educ.*, vol. 166, no. November 2020, p. 104169, 2021.
- [55] K. Okoye, A. R. H. Tawil, U. Naeem, R. Bashroush, and E. Lamine, "A semantic rule-based approach supported by process mining for personalised adaptive learning," *Procedia Comput. Sci.*, vol. 37, no. December, pp. 203–210, 2014.
- [56] R. Rajendran, A. Munshi, M. Emara, and G. Biswas, "A temporal model of learner behaviors in OELEs using process mining," in *ICCE 2018 - 26th International Conference on Computers in Education, Main Conference Proceedings*, 2018, no. December, pp. 276–285.
- [57] H. Jeong, G. Biswas, J. Johnson, and L. Howard, "Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models," *Educ. Data Min. 2010 - 3rd Int. Conf. Educ. Data Min.*, no. October, pp. 81–90, 2010.
- [58] K. Engelmann and M. Bannert, "Analyzing temporal data for understanding the learning process induced by metacognitive prompts," *Learn. Instr.*, vol. 72, no. October 2018, p. 101205, 2021.
- [59] J. M. E. M. Van Der Werf, C. Van Schuppen, S. Brinkkemper, S. Jansen, P. Boon, and G. Van Der Plas, "Architectural intelligence: A framework and application to e-learning," in *CEUR Workshop Proceedings*, 2017, vol. 1859, pp. 95–102.
- [60] G. Sedrakyan, M. Snoeck, and J. De Weerd, "Process mining analysis of conceptual modeling behavior of novices - Empirical study using JMermaid modeling and experimental logging environment," *Comput. Human Behav.*, vol. 41, pp. 486–503, 2014.
- [61] T. Doleck, R. B. Basnet, E. G. Poitras, and S. P. Lajoie, "Mining learner–system interaction data: implications for modeling learner behaviors and improving overlay models," *J. Comput. Educ.*, vol. 2, no. 4, pp. 421–447, 2015.
- [62] G. Deeva and J. De Weerd, "Understanding Automated Feedback in Learning Processes by Mining Local Patterns," in *BPM 2018 International Workshops. Lecture Notes in Business Information Processing*, vol. 342, 2019, pp. 56–68.
- [63] W. Poncin, A. Serebrenik, and M. Van Den Brand, "Mining student capstone projects with FRASR and ProM," in *SPLASH'11 Compilation - Proceedings of OOPSLA'11, Onward! 2011, GPCE'11, DLS'11, and SPLASH'11 Companion*, 2011, pp. 87–95.
- [64] A. Neyem, J. Diaz-Mosquera, J. Munoz-Gama, and J. Navon, "Understanding Student Interactions in Capstone Courses to Improve Learning Experiences CSS Concepts • Social and professional topics → Project and people management • Software and its engineering → Software creation and

- management General Terms,” in *SIGCSE 2017 - Proceedings of the 48th ACM Technical Symposium on Computer Science Education*, 2017, pp. 423–428.
- [65] J. Chen, Y. Zhang, J. Sun, Y. Chen, F. Lin, and Q. Jin, “Personalized micro-learning support based on process mining,” *Proc. - 2015 7th Int. Conf. Inf. Technol. Med. Educ. ITME 2015*, pp. 511–515, 2016.
- [66] M. Bannert, P. Reimann, and C. Sonnenberg, “Process mining techniques for analysing patterns and strategies in students’ self-regulated learning,” *Metacognition Learn.*, vol. 9, no. 2, pp. 161–185, 2014.
- [67] C. Sonnenberg and M. Bannert, “Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior?,” *Comput. Human Behav.*, vol. 96, pp. 259–272, 2019.
- [68] S. Estupiñán and N. Szilas, “When did i lose them? using process mining to study user engagement in interactive digital Narratives,” in *Interactive Storytelling. 12th International Conference on Interactive Digital Storytelling, ICIDS 2019*, 2019, vol. 11869 LNCS, pp. 374–378.
- [69] H. Nguyen, K. Y. Lim, L. L. Wu, C. Fischer, and M. Warschauer, “‘We’re looking good’: Social exchange and regulation temporality in collaborative design,” *Learn. Instr.*, vol. 74, no. October 2020, p. 101443, 2021.
- [70] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo, “Analytics of learning strategies: Associations with academic performance and feedback,” *ACM Int. Conf. Proceeding Ser.*, pp. 461–470, 2019.
- [71] J. Saint, D. Gašević, W. Matcha, N. A. A. Uzir, and A. Pardo, “Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning,” in *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK’20)*, 2020, pp. 402–411.
- [72] R. Martinez-Maldonado, K. Yacef, and J. Kay, “Data mining in the classroom: Discovering groups’ strategies at a multi-tabletop environment,” in *Proceedings of the 6th International Conference on Educational Data Mining, EDM 2013*, 2013, no. July 2015.
- [73] J. Saint, D. Gašević, and A. Pardo, “Detecting Learning Strategies Through Process Mining,” in *Lifelong Technology-Enhanced Learning. EC-TEL 2018. Lecture Notes in Computer Science*, 2018, vol. 11082 LNCS, pp. 385–398.
- [74] M. C. Rodriguez, M. L. Nistal, F. A. M. Fonte, M. L. Penin, and M. M. Molina, “Exploring the application of process mining to support self-regulated learning: An initial analysis with video lectures,” in *IEEE Global Engineering Education Conference, EDUCON*, 2018, vol. 2018-April, no. August 2019, pp. 1766–1774.
- [75] A. Bolt, M. de Leoni, W. M. P. van der Aalst, and P. Gorissen, “Business process reporting using process mining, analytic workflows and process cubes: A case study in education,” in *SIMPDA 2015*, 2017, vol. 244, pp. 28–53.
- [76] W. M. van der Aalst, S. Guo, and P. Gorissen, “Comparative Process Mining in Education: An Approach Based on Process Cubes,” in *International Symposium on Data-Driven Process Discovery and Analysis*, 2013, vol. 203, pp. 110–134.
- [77] S. Heirweg, M. De Smul, E. Merchie, and G. Devos, *Mine the process: investigating the cyclical nature of upper primary school students’ self-regulated learning*, no. 48. Springer Netherlands, 2020.
- [78] R. Lira *et al.*, “Process-oriented feedback through process mining for surgical procedures in medical training: The ultrasound-guided central venous catheter placement case †,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 11, pp. 1–21, 2019.
- [79] Y. Dahari, A. Gal, and A. Senderovich, “EDU-ProM: ProM for the classroom,” in *CEUR Workshop Proceedings*, 2017, vol. 1920.
- [80] F. Babič, J. Paralič, P. Bednár, and M. Raček, “Analytical framework for mirroring and reflection of user activities in E-learning environment,” *Adv. Intell. Soft Comput.*, vol. 80, no. September, pp. 287–296, 2010.
- [81] C. M. Chen and W. F. Wang, “Mining Effective Learning Behaviors in a Web-Based Inquiry Science Environment,” *J. Sci. Educ. Technol.*, vol. 29, no. 4, pp. 519–535, 2020.
- [82] W. Premchaiswadi and P. Porouhan, “Process modeling and decision mining in a collaborative distance learning environment,” *Decis. Anal.*, vol. 2, no. 1, 2015.
- [83] W. Peeters, M. Saqr, and O. Viberg, “Applying learning analytics to map students’ self-regulated learning tactics in an academic writing course,” *ICCE 2020 - 28th Int. Conf. Comput. Educ. Proc.*, vol. 1, no. September, pp. 245–254, 2020.

- [84] Agusriandi, I. S. Sitanggang, and S. H. Wijaya, "Student Performance Based on Activity Log on Social Network and e-Learning," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, pp. 2276–2281, 2020.
- [85] M. Mittal and A. Sureka, "Process mining software repositories from student projects in an undergraduate software engineering course," in *36th International Conference on Software Engineering, ICSE Companion 2014 - Proceedings*, 2014, pp. 344–353.
- [86] J. A. Caballero-Hernández, A. Balderas, M. Palomo-Duarte, P. Delatorre, A. J. Reinoso, and J. M. Dodero, "Teamwork assessment in collaborative projects through process mining techniques," *Int. J. Eng. Educ.*, vol. 36, no. 1B, pp. 470–482, 2020.
- [87] J. A. C. Hernández, M. P. Duarte, and J. M. Dodero, "An architecture for skill assessment in serious games based on Event Sequence Analysis," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1322, 2017.
- [88] J. A. Ruipérez-Valiente, L. Rosenheck, and Y. J. Kim, "What Does Exploration Look Like? Painting a Picture of Learning Pathways Using Learning Analytics," in *Game-Based Assessment Revisited*, no. November, 2019, pp. 281–300.
- [89] T. Schaedler Uhlmann, E. Alves Portela Santos, and L. A. Mendes, "Process Mining Applied to Player Interaction and Decision Taking Analysis in Educational Remote Games," in *Smart Industry & Smart Education. REV 2018*, 2019, vol. 47, pp. 425–434.
- [90] S. A. Priyambada, E. R. Mahendrawathi, and B. N. Yahya, "Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering," in *4th Information Systems International Conference 2017, ISICO 2017*, 2017, vol. 124, pp. 264–273.
- [91] L. Y. Bendatu and B. N. Yahya, "Sequence Matching Analysis for Curriculum Development," *J. Tek. Ind.*, vol. 17, no. 1, pp. 47–52, 2015.
- [92] N. Trčka and M. Pechenizkiy, "From local patterns to global models: Towards domain driven educational process mining," in *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*, 2009, pp. 1114–1119.
- [93] M. Cameranesi, C. Diamantini, L. Genga, and D. Potena, "Students' Careers Analysis: a Process Mining Approach," in *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics (WIMS '17)*, 2017, pp. 1–7.
- [94] J. P. Salazar-Fernandez, M. Sepúlveda, J. Munoz-Gama, and M. Nussbaum, "Curricular analytics to characterize educational trajectories in high-failure rate courses that lead to late dropout," *Appl. Sci.*, vol. 11, no. 4, pp. 1–15, 2021.
- [95] J. P. Salazar-Fernandez, M. Sepúlveda, and J. Munoz-Gama, "Describing educational trajectories of engineering students in individual high-failure rate courses that lead to late dropout," in *CEUR Workshop Proceedings*, 2019, vol. 2425, pp. 39–48.
- [96] J. A. Caballero-Hernandez, J. M. Dodero, I. Ruiz-Rube, M. Palomo-Duarte, J. F. Argudo, and J. J. Dominguez-Jimenez, "Discovering bottlenecks in a computer science degree through process mining techniques," in *SIIE 2018 - 2018 International Symposium on Computers in Education, Proceedings*, 2018, no. 1, pp. 1–6.
- [97] V. Aisa, A. P. Kurniati, and A. W. Yanuar Firdaus, "Evaluation of the online assessment test using process mining (Case Study: Intensive English Center)," in *2015 3rd International Conference on Information and Communication Technology, ICoICT 2015*, 2015, pp. 472–477.
- [98] A. Piotrkowicz, V. Dimitrova, and T. E. Roberts, "Temporal Analytics of Workplace-Based Assessment Data to Support Self-regulated Learning," in *EC-TEL 2018: 13th European Conference on Technology Enhanced Learning. Lecture Notes in Computer Science*, 2018, vol. 11082 LNCS, pp. 570–574.
- [99] A. H. Cairns *et al.*, "Using semantic lifting for improving educational process models discovery and analysis," in *CEUR Workshop Proceedings*, 2014, vol. 1293, pp. 150–161.
- [100] A. Hicheur Cairns, B. Gueni, H. Hafdi, C. Joubert, and N. Khelifa, "Towards a distributed computation platform tailored for educational process discovery and analysis," in *International Conference on Protocol Engineering, ICPE 2015 and International Conference on New Technologies of Distributed Systems, NTDS 2015 - Proceedings*, 2015, pp. 1–8.
- [101] J. Munoz-Gama, V. Gálvez, R. de la F. Sanhueza, M. Sepulveda, and R. S. Fuentes, "Interactive Process Mining for Medical Training," in *Interactive Process Mining in Healthcare*, C. Fernandez-Llata, Ed. Springer, Cham, 2021, pp. 233–242.
- [102] A. Sypsas and D. Kalles, "Computing Similarities Between Virtual Laboratory Experiments Models using Petri Nets," in *20th International Conference on Modelling and Applied Simulation, MAS 2021*, 2021, pp. 29–37.

- [103]D. Riofrío-Luzcando and J. Ramírez, “Predictive student action model for procedural training in 3D virtual environments,” in *E-Learning, E-Education, and Online Training. Second International Conference, eLEOT 2015*, 2016, pp. 37–62.
- [104]J. C. Vidal, M. Lama, B. Vázquez, and M. Mucientes, “Reconstructing IMS LD units of learning from event logs,” in *Proceedings of the 9th European Conference on Open Learning and Teaching in Educational Communities*, 2014, vol. 8719 LNCS, pp. 345–358.
- [105]J. C. Vidal, B. Vázquez-Barreiros, M. Lama, and M. Mucientes, “Recompiling learning processes from event logs,” *Knowledge-Based Syst.*, vol. 100, no. C (May 2016), pp. 160–174, 2016.
- [106]L. Castillo, “A virtual laboratory for multiagent systems: Joining efficacy, learning analytics and student satisfaction,” in *2016 International Symposium on Computers in Education, SIIE 2016: Learning Analytics Technologies*, 2016, pp. 1–6.
- [107]B. M. Toussaint and V. Luengo, “Mining surgery phase-related sequential rules from vertebroplasty simulations traces,” in *15th International Conference on Artificial Intelligence in Medicine*, 2015, vol. 9105, pp. 35–46.
- [108]B. Fernández-Gallego, M. Lama, J. C. Vidal, and M. Mucientes, “Learning analytics framework for educational virtual worlds,” in *International Conference on Virtual and Augmented Reality in Education*, 2013, vol. 25, pp. 443–447.
- [109]R. Dijkman and A. Wilbik, “Linguistic summarization of event logs – A practical approach,” *Inf. Syst.*, vol. 67, pp. 114–125, 2017.
- [110]W. Van der Aalst, *Process mining: Data science in action*, Second Edi. 2016.

AUTHORS

Athanasios Sypsas is a PhD candidate at Hellenic Open University, School of Science and Technology, Greece. His main scientific interests are Simulation, Artificial Intelligence, Distance Learning and Programming.



Dimitris Kalles is a Professor on “Artificial Intelligence – Applications” with the Hellenic Open University (HOU) and Director of the undergraduate study programme “Computer Science”. His main research interests are in Artificial Intelligence and Machine Learning and his work also has a strong focus in Educational Technology and Software Engineering. He has published over 100 papers in scientific journals and conference proceedings and his work has received more than 1000 citations.

