Overview

A survey on educational process mining

Alejandro Bogarín,[1] Rebeca Cerezo[2] and Cristóbal Romero[1]*

Educational process mining (EPM) is an emerging field in educational data mining (EDM) aiming to make unexpressed knowledge explicit and to facilitate better understanding of the educational process. EPM uses log data gathered specifically from educational environments in order to discover, analyze, and provide a visual representation of the complete educational process. This paper introduces EPM and elaborates on some of the potential of this technology in the educational domain. It also describes some other relevant, related areas such as intentional mining, sequential pattern mining and graph mining. It highlights the components of an EPM framework and it describes the different challenges when handling event logs and other generic issues. It describes the data, tools, techniques and models used in EPM. In addition, the main work in this area is described and grouped by educational application domains.

How to cite this article:
WIREs Data Mining Knowl Discov 2018, 8:e1230. doi: 10.1002/widm.1230

INTRODUCTION

Nowadays, with the development and increasing popularity of technology supported learning environments, information systems enable us to capture all student events/actions/activities at different levels of granularity, from low level events such as keystrokes, mouse gestures and clicks, to higher-level events such as students' learning activities.[1] These systems have tracking and logging capabilities to gather different types of temporal data such as click streams, chat logs, document edit histories (e.g., wikis, etherpads), motion tracking (e.g., eye-tracking, Microsoft Kinect), learning resource usage logs, and various interaction logs (e.g., with intelligent tutoring systems). Process mining (PM) can use these so-called logs, audit trails, log files or traces in order to discover, monitor and improve educational processes. PM provides a bridge between data mining (DM) and process modeling and analysis. PM as a sub-discipline of DM adds the process-oriented view of the DM procedure.[2] DM has been widely applied successfully to find interesting patterns from data gathered in educational environments.[3] However, educational data mining (EDM) techniques focus on data dependencies or simple patterns and do not provide a visual representation of the overall learning process. EDM does not focus on the process as a whole.[4] Classical DM techniques such as classification, clustering, regression, association rule mining (ASR) and sequence mining are of little use for control-flow discovery and are not process-centric. To allow for these types of general analysis, in which the process rather than the result plays the central role, a new method of data-mining research, called PM, has been proposed.[1] More specifically, educational process mining (EPM) is the application of PM to raw educational data.[5] Both EDM and EPM apply specific algorithms to data in order to uncover hidden patterns and relationships. But EDM techniques are not process-centric and do not focus on event data.[2] For EDM techniques the rows (instances) and columns (variables) of a typical data file do not have any meaning. On the other hand, EPM is process-centric thereby making the unknown (or partially known) processes explicit; it assumes a different type of data: events. Each event belongs to a single process instance (also called a case), and events are related to activities. EPM is interested in end-to-end

*Correspondence to: cromero@uco.es

[1]Department of Computer Science, University of Cordoba, Cordoba, Spain

[2]Department of Psychology, University of Oviedo, Oviedo, Spain

Conflict of interest: The authors have declared no conflicts of interest for this article.

processes rather than local patterns. End-to-end process models and concurrency are essential for PM.[6] Furthermore, other suitable approaches with the same scope such as process discovery, conformance checking, and bottleneck analysis, are not addressed by traditional EDM techniques.

On this basis, this paper is arranged as follows: *Background and Related Areas* in second section. *Framework and Concepts* introduces the EPM framework. *Data and Tools* describes the type of data, formats and tools commonly used in EPM. *Techniques* section, outlines the most commonly used techniques. *Application Domain* describes the different application domains in education and the most relevant work. Finally, *Conclusions and Further Research* are outlined.

BACKGROUND AND RELATED AREAS

PM is a relatively new technology emerging from the business community.[2] It focuses on the development of techniques aimed at extracting process-related knowledge from event logs. It uses event logs recorded by information systems in order to discover, monitor and improve processes in different domains as well as to check process conformance, detect bottlenecks, and predict execution problems. PM is also known as Workflow Mining (WM). Most work in PM has concentrated on (business) workflow systems and the discovery of Petri nets representations of workflows.[7] The methods described by these labels take information from event logs as input and produce process models that describe the information in the event logs in a comprehensive manner.[8] The process-oriented view helps to exceed the mainly isolated view of datasets that dominates traditional DM.

PM provides new means of improving processes in a variety of domains. In our case, process-oriented knowledge discovery techniques in educational systems are an upcoming and emerging field of interest. PM applied to education data is called EPM. The combination of learning technology with PM brings considerable potential.[5] EPM involves the discovery, analysis and enhancement of processes and flows underlying the event logs generated by virtual learning environments (VLEs). However, we should bear in mind that there are limits to the value of an inductive, data-driven approach if there is no theory that can guide mining. Depending on the kind of process measures considered, students' navigation behavior could be relatively easily mined from a technical point of view. However, it should be used to assure initial efforts towards a guiding concept, rather than to start looking for regularities and patterns without a guiding conceptual framework. Taking this into consideration, EPM is able to build complete, compact educational process models that are able to reproduce all the observed behaviors, check to see if the modeling behavior matches the behavior observed, and project extracted information from the registrations in the pattern to make the tacit knowledge explicit, and to facilitate a better understanding of the process.[7] Likewise, the term Learnflow Mining in correspondence with Workflow Mining has been used by some authors[9,10] whereas many more authors[1,11–13] prefer the term EPM in correspondence with PM, which is the currently the most commonly used term.

There are also other related research areas used to discover learners' behavior (see Table 1). Next, we briefly address three of those which are most closely related to PM: intention mining (IM), sequential pattern mining (SPM), and graph mining (GM).

TABLE 1 | Main Related Areas with EPM

	Objectives	Algorithms	Models	Tools
Process mining	Discover underlying processes from event logs	Heuristic Miner, Fuzzy Miner, etc.	Petri Nets, Heuristic Net, BMMN, etc.	ProM, Disco, Celonis, etc.
Intention mining	Model the processes according to the purpose of the actors	Viterbi Algorithm, Baum-Welch Algorithm, etc.	KAOS, I*, Map, etc.	No tools found
Sequence pattern mining	Find common patterns between data examples where the values are delivered in a sequence	Generalized Sequential Patterns (GSP), Sequential Pattern Mining (SPAM), PrefixSpan, etc.	Sequences and subsequences, rules	SPFM, Himalaya Data Mining, etc.
Graph mining	Extract patters (sub-graphs) of interest from graphs that describe the underlying data	Branch-and-bound, On-line Plan Recognition, Recursive Matrix (R-MAT), etc.	Probabilistic graphs, signed graphs, colored graphs, Transition graphs, etc.	Graphviz, Deep Though, GSLAP, etc.

Intention Mining

IM or Intentional process mining is another recently emerged, related field of research. This field has the same objectives as PM but specifically addresses intentional process models, that is, processes focused on the reasoning behind activities.[14]

It is important to note that we have not found any research about the application of IM in education but the potential of this technique can be easily understood if it is particularly suitable to users' ways of thinking and working as it captures the human reasoning behind activities.

Sequence Pattern Mining

SPM[15] is a very commonly used technique in the DM environment for discovering common sub-sequences. Sequential pattern analysis aims to find relationships between occurrences of sequential events, that is, to find if any specific order of occurrences exists.[16] SPM is related to episode mining (EP); in fact, both techniques can be seen as variants of ASR. However, SPM methods find the most frequent event patterns across a set of event sequences, while EP discovers the most frequently used event patterns within a given sequence. There are other SPM-related techniques such as lag sequential analysis (LAS), t-pattern analysis and Markov models. While t-pattern analysis can be used to explore longer, more temporally separated sequences than LAS and Markov models, all these techniques are best suited to relatively short recurring sequences and analysis of event transitions.[17]

SPM techniques have been widely applied to analyze student learning behaviors. But, they are more indicated when trying to discover serial or simpler behavioral patterns than a process, for instance, learner activity paths in a course. So, SPM is not appropriate for discovering learning behaviors that describe the overall learning process.[18]

Graph Mining

GM is another popular pattern mining technique. The goal of GM or sub-graph mining is to find all frequent sub-graphs in a large graph or a database of graphs. GM and DM are closely related. Nevertheless, the main goal of graph mining is to supply new ideas and efficient algorithms for mining topological substructures embedded in graph data, while the main goal of DM is to supply ideas for mining and/or learning relational patterns represented by expressive logical languages. The former is more geometry-oriented and the latter more logic and relation oriented.[19] It is also important to differentiate between GM and social network analysis (SNA); SNA can be considered as an application of GM.

Graph-based Educational Data Mining (G-EDM) is a new, related, research area. Both G-EDM and EPM use graphs to represent information. However, while the task of GM is to extract patterns (sub-graphs of interest) from graphs that describe the underlying data and could be used further, for example, for classification or clustering, PM focusses on the process as a whole and therefore its graphs discover the overall learning process. Graphs are extremely important in the EDM community because of many types of data can be represented naturally as graphs including social network data and online discussions.

Finally, Table 1 shows a comparison of the previously described EPM related research areas.

FRAMEWORK AND CONCEPTS

An overview of the application of PM in the educational field is shown in Figure 1. In previous sections we have addressed how educational processes can be translated into executable processes by means of models. This EPM framework is an adaptation of the generic framework of PM[20] to the field of education[13,21] and below we describe the main components:

- **Educational world**: Basically two participants play an important role in any e-learning activity, teachers and students. Teachers supply appropriate resources to ensure student success. Students are the essential part of any e-learning activity, interacting with other participants (students or teachers), and with the system itself. Finally, courses, lectures, exams, etc. are used as resources for participants.
- **Virtual learning environment**: This supplies the basic structures and resources where the participants' learning actions and interactions occur. It also logs the events that occur during the e-learning process. Most provide teachers or researchers with basic tools for analyzing students' learning (marks evolution, number of activities done, forum participation, last log in, etc.) but do not provide specific tools that would allow educators to thoroughly assess the overall student learning process.
- **Event logs**: These are files that record events that occur in VLEs and are normally stored in databases. They contain a large amount of raw data about the educational agents' interaction with the VLE. Event logs need to be

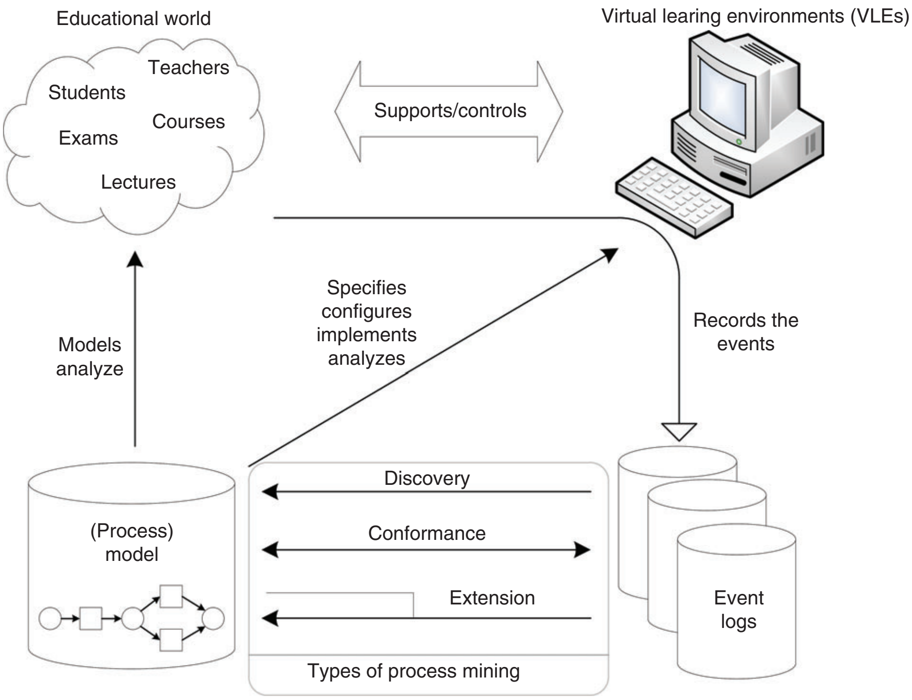

FIGURE 1 | EPM framework: types and components.

transformed into a particular file format in order to can be used by a specific PM tool.

- **Process models:** These uncover valuable information about how the participants of the educational world interact with the system starting from the event logs. They are obtained using different techniques in order to discover processes relevant to learning. Three main types of PM (see Table 2) can be distinguished[20]: discovery, conformance, and extension. These three basic types of PM can be also explained in terms of input and output (see Figure 2).

In addition to the three main types of PM, PM also provides distinct perspectives[22] such as control-flow, organizational, case, and time perspectives. The most commonly used in the educational environment is the control-flow perspective that focuses on the ordering of activities. The principal aim of this perspective is to discover an ideal description of all imaginable learning paths or trails in education[23] that can be generated when students navigate through a learning environment.

DATA AND TOOLS

In this section, we provide a deeper description of the data, potential difficulties, and software solutions used to perform EPM analysis.

TABLE 2 | Types of Process Mining

Type	Description	Application in Education
Process discovery	Constructs a comprehensive process model able to reproduce the behavior seen in the log file.	Instructor can visualize the behavior model of students' learning paths providing knowledge of the process instead of only the learning result
Conformance checking	Finds deviations between observed behaviors in event logs and generated process models.	Instructor can analyze whether the model corresponds to the behavior model in event logs and, for instance, find outliers.
Extension or enhancement	Aims to improve or extend a given process model based on information extracted from a specific event log related to the same process.	Instructor can detect bottlenecks or relationships between students in a course because different approaches can merged from a single integrated and extended process model.

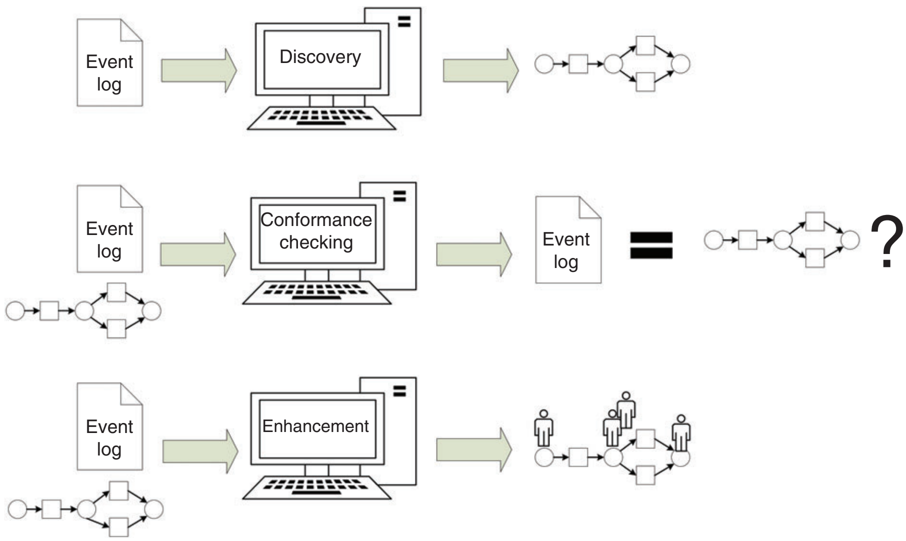

FIGURE 2 | Types of process mining explained in terms of input and output.

The starting point for PM is an event log.[22] An event log may be an Excel spreadsheet, a database table or a simple file that contains a trace/sequence of events. Each event is a row in the event log and refers to a case (case id), an activity (activity name), and a point in time (timestamp), sometimes it may contain additional information. Generally, they need to be transformed into specific formats for storing logs such as XES (eXtensible Event Stream) or MXML (Mining eXtensible markup language) in order to be used by a PM tool.[5] There are some specific tools supporting the conversion of different data sources to these formats such as ProMimport.[22]

Educational event logs can be gathered from a wide range of virtual e-learning environments such as Learning Management Systems (LMSs), Massive Open Online Courses (MOOCs), intelligent Tutoring Systems (ITSs), and Adaptive Hypermedia Systems (AHSs). See Figure 3 for an example of an event log generated from Moodle LMS. The Moodle system

Time	IP address	Full name	Action	Information
24/09/2013 12:22	150.214.10.12	Student53	Resource view	Tema 1
24/09/2013 12:24	150.214.10.12	Student53	Resource view	Tema 2
24/09/2013 12:25	150.214.10.12	Student53	Resource view	Tema 3
24/09/2013 12:26	150.214.10.12	Student53	Resource view	Tema 4
24/09/2013 12:28	150.214.10.12	Student53	Resource view	Clases précticas (Diagnóstico)
24/09/2013 12:30	180.45.67.44	Student42	Resource view	Tema 1
24/09/2013 12:31	180.45.67.44	Student42	Resource view	Tema 2
24/09/2013 12:31	180.45.67.44	Student42	Resource view	Tema 3
24/09/2013 13:29	155.123.23.14	Student35	Folder view	Clases précticas (Diagnóstico)
24/09/2013 13:29	155.123.23.14	Student35	Resource view	Tema 1
24/09/2013 13:29	155.123.23.14	Student35	Resource view	Tema 2
24/09/2013 13:29	155.123.23.14	Student35	Resource view	Tema 3
24/09/2013 13:29	155.123.23.14	Student35	Resource view	Tema 4
24/09/2013 14:06	145.124.25.65	Student49	Folder view	Clases précticas (Diagnóstico)
24/09/2013 14:33	154.132.45.66	Student7	Folder view	Clases précticas (Diagnóstico)
24/09/2013 14:33	154.132.45.66	Student7	Resource view	Tema 1
24/09/2013 14:33	154.132.45.66	Student7	Resource view	Tema 2
24/09/2013 14:33	154.132.45.66	Student7	Resource view	Tema 3
24/09/2013 14:33	154.132.45.66	Student7	Resource view	Tema 4

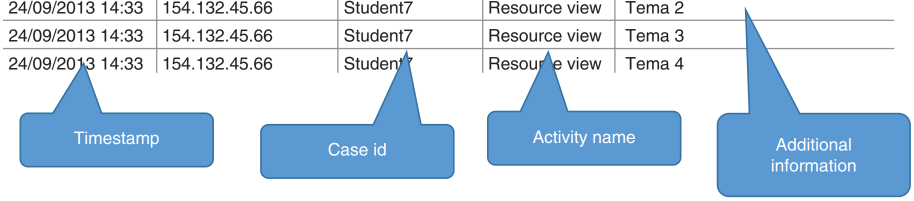

FIGURE 3 | Example of moodle event log.

logs every click (Time, IP, Student-FullName, Action and Additional-Information) that educational agents make for navigational purposes, generating a vast amount of *a priori* senseless information.

In general, several hurdles appear when handling event logs and they need to be overcome and borne in mind in EPM;[13,22] see Table 3 for further information.

Finally, many tools have emerged to support PM techniques such as:[22] ProM, Disco, Celonis Discovery, Perceptive Process Mining, QPR ProcessAnalyzer, Aris Business Process Analysis, Fujitsu Process Analytics, XMAnalyzer, and StereoLOGIC Discovery Analyst. However, all of them are general purpose PM tools and only a few have been used in EPM (see Table 4).

Only three of these PM tools have been referenced in all of the papers (see Table 4). ProM tool is a generic open-source framework for implementing PM and it is the most complete and most commonly used in EPM, followed by Disco, which is also a general purpose tool but commercial. There is only one PM software specific to the education domain, named SoftLearn.[24] It provides a graphical interface that teachers can use to visualize learning paths as activity graphs and to access the relevant data generated in the learning activities.

TECHNIQUES

In this section, we describe the most commonly used techniques in EPM. We highlight four main groups

TABLE 3 | Challenges when handling event logs

Issue	Description	Example in EPM
Correlation	Events are grouped per case in an event log. Events need to be related to each other.	Students perform similar kinds of actions in a forum.
Noise	An event log may contain outliers. Exceptional behavior is not representative of typical behavior of the process.	Students can leave an open session.
Incompleteness	A common problem is that the event log contains too few events to be able to discover some of the underlying control-flow structures.	E-learning systems fall down.
Distribution	Data may be distributed over a variety of sources.	Student information can be gathered from different sources: Administrative information, theory and practice classroom, online learning environments, etc.
Timestamp	Events need to be ordered per case.	Typical problems: only dates, different time zones, delayed logging
Snapshots	Cases may have a lifetime extending beyond the recorded period	Student case was started before the beginning of the event log.
Scoping	What is the *process* we want to investigate? How to decide which tables to include?	LMS and MOOC can provide different tables to investigate different process.
Granularity	The events in the event log are at a different level of granularity.	Educational information may have different levels of granularity, ranging from low level clicks, to activities, courses, etc.
Contextualization	Events occur in a particular context. This context may explain certain phenomena. This requires the merging of event data with contextual data	Teachers discover models in a repeat student class.
Size	Number of cases or events in event logs can be high. These files can be difficult to handle due to their size.	Virtual learning environments can generate huge logs.
Complexity	Distinct traces and activities in event logs can be high due to the large diversity of behaviors in students' learning paths.	Educational environments can generate complex models that are difficult to understand, named spaghetti models.
Concept drift	Situation in which the process changes while being analyzed.	Courses and study curriculums may be modified at any time during learning span.
Privacy	Privacy and authentication has many ethical dimensions.	Students need to be aware what the system is doing with their data.

TABLE 4 | Comparison between the Main Tools Used in EPM

Company (Country)	**ProM** Eindhoven Technical University (Netherlands)	**Disco** Fluxicon (Netherlands)	**SoftLearn** University of Santiago de Compostela (Spain)
Purpose	General	General	Specific (education)
Type	Free	Commercial	Private
Filtering	Yes	Yes	No
Process discovery	Yes	Yes	Yes
Conformance checking	Yes	No	No
Social network mining	Yes	No	No
Number of papers/Works	21	7	1

of techniques: discovery, conformance checking, dotted chart analysis, and SNA.

Discovery Techniques

Process discovery techniques build a process model based solely on an event log by capturing the behavior seen in the log. They focus on the control-flow perspective of process. There are a lot of algorithms in PM for discovering underlying processes from event logs, but the most often used in educational domains are (see Table 5):

- **Alpha algorithm:** a relatively intuitive and simple technique based on dependency relation between events which requires ideal event logs without noise. It was one of the first algorithms that was able to deal with concurrency.[25]
- **Heuristic Miner algorithm:** this uses likelihood by calculating the frequencies of relations between the tasks (e.g., causal dependency, loops, etc.) and constructs dependency/frequency tables and dependency/frequency graphs.[14] The Heuristic Miner algorithm was designed to make use of a frequency based metric and so is less sensitive to noise and the incompleteness of logs.[26]
- **Genetic algorithm:** this provides process models built on causal matrixes (input and output dependencies for each activity). This approach tackles problems such as noise, incomplete data, non-free-choice constructs, hidden activities, concurrency, and duplicate activities.[14]
- **Fuzzy miner:** this is one of the newer process discovery algorithms. It is the first algorithm to directly address the problems of large numbers of activities and highly unstructured behavior.[27]

TABLE 5 | Representation Models Used in EPM Works

Work/Paper	Petri Nets	HLPN	Fuzzy	Heuristic
Weijters et al.[4]	X			X
Günther and Van Der Aalst[27]			X	
Pechenizkiy et al.[20]	X		X	X
Reimann et al.[17]	X			X
Trcka and Pechenizkiy[7]		X		
Southavilay et al.[39]				X
Trcka et al.[1]	X		X	
Poncin et al.[49]			X	
Schoor and Bannert[37]			X	
Anuwatvisit et al.[48]	X			
Ayutaya et al.[47]	X			X
Bergenthum et al.[9]	X	X		
van der Aalst et al.[12]			X	
Reimann et al.[8]			X	
Bannert et al.[18]	X		X	
Cairns et al.[40]				X
Cairns et al.[32]			X	
Bogarin et al.[26]				X
Cairns et al.[31]	X			X
Cairns et al.[13]			X	X
Mukala et al.[34]			X	
Ariouat et al., 2016[43]				X
Doleck et al.[41]			X	
Okoye et al.[55]			X	
Sedrakyan et al.[54]			X	
Vahdat et al.[42]			X	
Vidal et al.[21]	X			

Good notation is necessary in order to represent ready process models to the end-user. All the above mentioned algorithms produce a process model that is normally independent of the desired representation. There are different types of representations or metamodels in PM such as Petri nets, Workflow nets, Fuzzy nets, Heuristic nets, Causal nets, Process tree, BPMN (Business Process Model and Notation), EPC (Event Driven Process Chain), and UML (Unified Modeling Language) Activity Diagram. Although Petri nets and BPMN are the most often used in PM,[14] the most commonly used in education domain are (see Table 5):

- **Petri Nets:** Graphs with two types of nodes linked by directed arcs. The first type of node is known as place and is represented by an ellipse. Places can store a multi-set of values, called tokens. Transitions are represented as rectangles and identify active elements of the net.[28]
- **High-level Petri Net (HLPN):** Extended classical Petri Nets with color, time and hierarchy. Colored Petri Nets (CPN) were the first concrete realization of HLPN and were a graphical language for analyzing the properties of concurrent systems.[14]
- **Fuzzy net:** Simplifies the complete model by preserving highly significant events or edges, aggregating less significant but highly correlated edges and nodes by clustering, and abstracting from less significant and poorly correlated edges and nodes by removing them from the simplified model.[27]
- **Heuristic nets:** A directed cycle graph which represents the most frequent behaviors of the students in the dataset used. In heuristic nets the square boxes represent the actions and the arcs/links represent dependences/relations between actions.[26]

Additionally, it is possible to automatically transform a model from one representation/notation to another when using some PM software. In Figure 4, we show two different representations/notations obtained from the same log file. A Petri Net showing the causality and parallelism of the events and a Heuristic Net showing the frequency of the events and how strong the dependency between events is.

It is necessary to state that the most commonly used representation model in EPM research is the Fuzzy net, followed by Petri net and Heuristic net, with HLPN the least used (see Table 5).

Conformance Checking Techniques

The goal of conformance checking is to find commonalities and discrepancies between the modeled behavior and the observed behavior. In the EPM literature, two techniques stand out in conformance checking:

- Linear Temporal Logic (LTL) Checker. This checks whether the logs satisfy some Linear Temporal Logic (LTL) formula.[29] LTL Checker does not compare a model with the log, but a

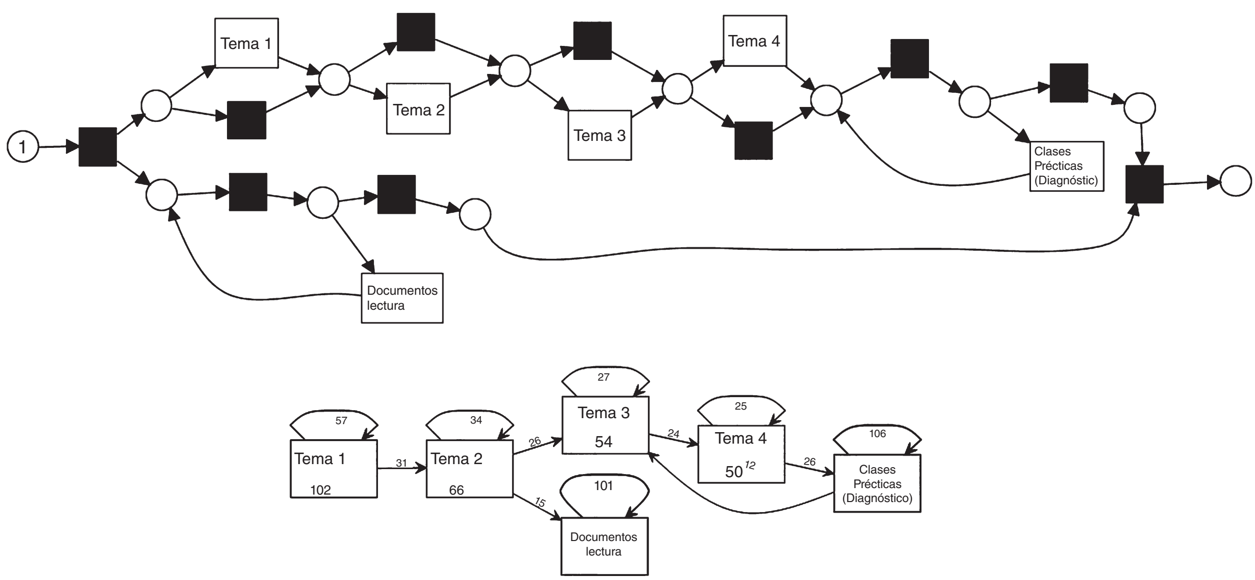

FIGURE 4 | Examples of Petri and Heuristic Net generated with the same log data.

set of requirements described by the (linear) temporal logic LTL.
- The conformance checker. This requires a model in addition to an event log. It replays an event log within a Petri net model in a non-blocking way while gathering diagnostic information that can be accessed afterwards.[30]

Dotted Chart Analysis Technique

A dotted chart shows the spread of events over time by plotting a dot for each event in an event log thus providing some insight into the underlying process, its performance and any patterns of interest. It represents the log file visually, showing a general time perspective of the process. The chart has two orthogonal dimensions: time and component types. The time is measured along the horizontal axis of the chart. The component types are shown along the vertical axis.[31] Figure 5 shows an example of dotted chart about the daily work carried out by students in Moodle. Each row is a different task or Moodle event in the course, and the size of dots represent how many students have done this task at a particular time.

Social Network Analysis Technique

SNA refers to the collection of methods, techniques and tools in sociometry aimed at the analysis of social networks. SNA aims to extract social networks from event logs based on the observed interactions between performers, depending on how process instances are routed between these performers.[32] A social network consists of nodes representing organizational entities and arcs representing relationships. Figure 6 shows an example of social networks that represents how and how much students interact in a Moodle forum. Bigger nodes represent more active students and Arcs represent the moment when they interact.

Finally, a summary of the most commonly used techniques in EPM research.

As Table 6 shows, the most commonly used discovery algorithms are Heuristic Miner and Fuzzy Miner. Conformance checker is the most commonly used conformance technique. The Dotted Chart is more commonly used in educational research than SNA.

APPLICATION DOMAINS

EPM has been used in a wide range of application domains in education in order to address varying educational problems. In this section, we describe the main body of literature, giving special importance to current EPM applications rather than the specific results.

MOOCs, LMS and Hypermedia Environments

Massive Open Online Courses (MOOCs), Learning Management System (LMS), Hypermedia and other similar online learning environments supply free learning opportunities to the online community. Log files generated by these systems provide an insight into how people follow the course, when they watch videos or lectures, and when they hand in tasks, among others.

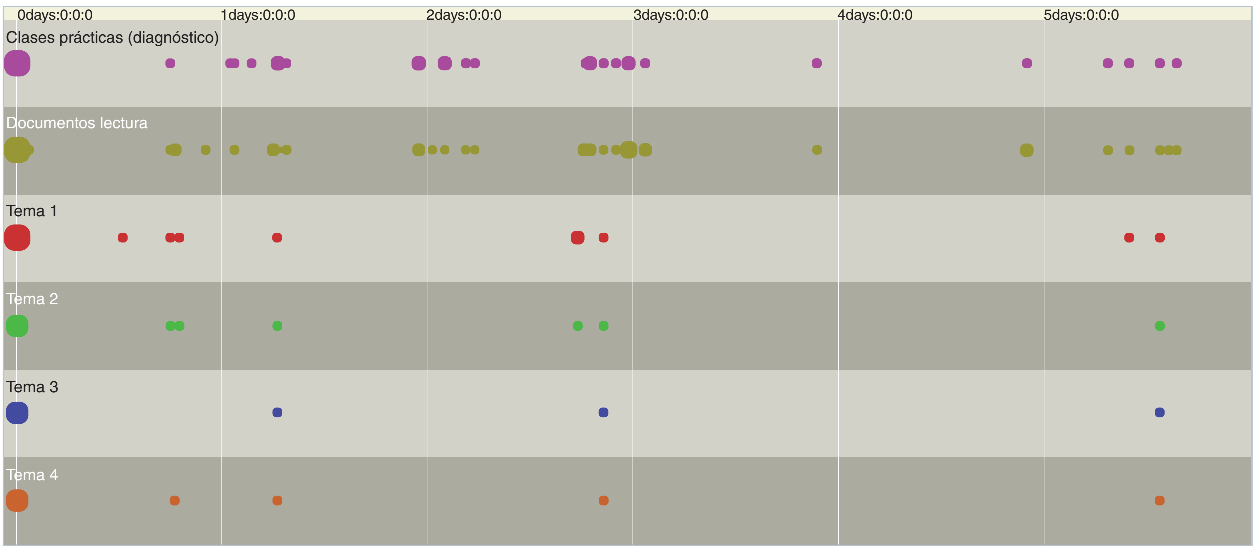

FIGURE 5 | Example of a dotted chart of the daily work carried out by students in Moodle.

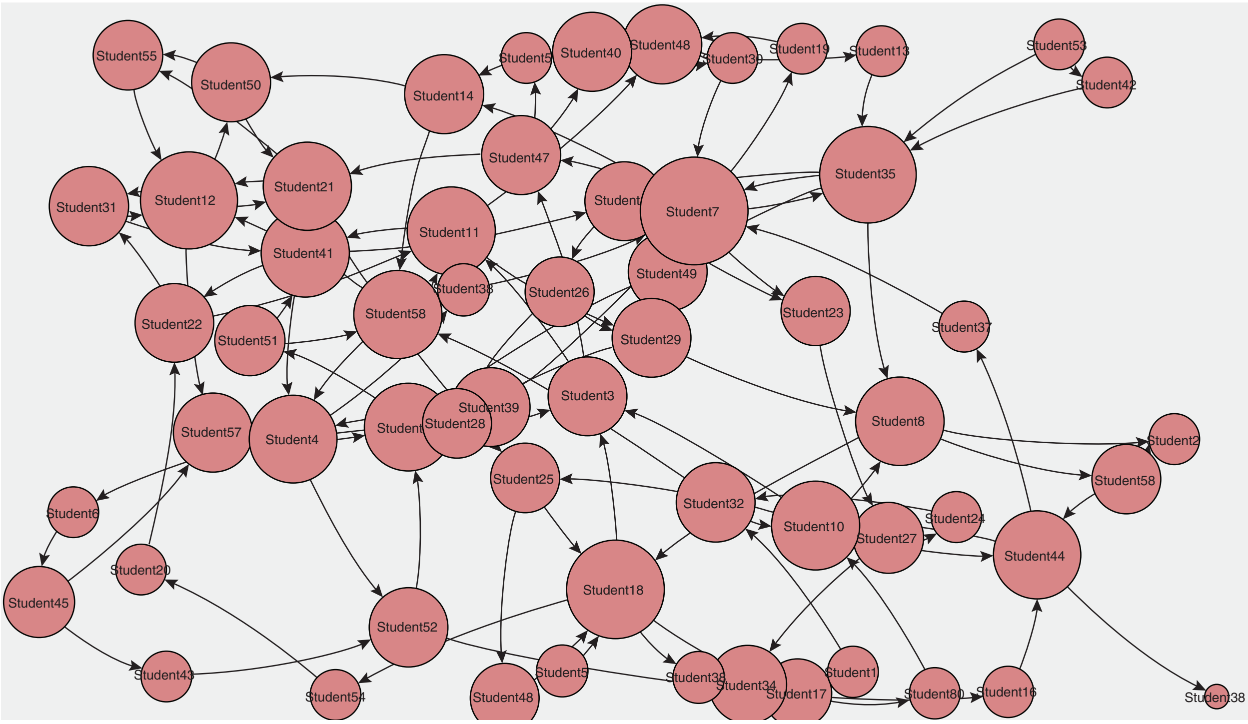

FIGURE 6 | Example of a social network that represents how and how much students interact in a Moodle forum.

There is a lot of research about the application of PM in this type of learning environment. Trcka et al.[1] exemplified the applicability of PM and discussed some of its potential for extracting knowledge from LMSs, considering only students´ examination traces. In Bogarin et al.,[26] the authors used data from Moodle logs and proposed using clustering in order to be able to obtain more specific and accurate Process Models of students' behavior. In a similar environment Reiman et al.[8] proposed the use of traces to study self-regulated learning (SRL) in a hypermedia environment based on theoretical principles and PM. Using those principles, Bannert et al.[18] detected differences in frequencies of SRL events using PM techniques and found that successful students demonstrate more learning and regulation events. In other research Mukala et al.[33] used PM techniques in order to trace and analyze students' learning habits based on MOOC data. Results indicated that successful students follow a sequentially-structured watching pattern while unsuccessful students are unpredictable and have poorly structured processes. In later research Mukala et al.[34] made use of alignment-based conformance checking to extract and analyze students' learning patterns in a MOOC. Along similar lines, Emond and Buffett[35] applied process discovery mining and sequence classification mining techniques to model and support SRL in heterogeneous environments of learning content, activities, and social networks. They used a data set of semi-structured learning activities taken from the DataShop at the Pittsburgh Science of Learning Centre. Finally, Vidal et al.[21] used logs from a VLE to extract the learning flow structure using PM, and to obtain the underlying rules that control students' adaptive learning by means of decision tree learning.

Computer-Supported Collaborative Learning

Computer-supported collaborative learning (CSCL) is characterized by the sharing and construction of knowledge between participants using technology as their primary means of communication or as a common resource.

PM has been applied in CSCL in order to supply feedback to students on their decision-making processes.[17] The goal was to use PM to identify models of decision-making groups that took place in a chat room. In a similar study, Bergenthum et al.[9] proposed a modeling language for collaborative learnflows that specifically takes the actors, the roles, and the explicit representation of groups into account. Their research extends on previous work focused on the discovery of the structures for flow control using methods from the area of WM, with

TABLE 6 | Techniques used in EPM research

Work/Paper	Discovery Algorithm	Conformance Techniques	Dotted Chart	SNA
Weijters et al.[4]	Heuristic Miner			
Pechenizkiy et al.[20]	Heuristic Miner, Fuzzy Miner	Conformance checker	X	
Reimann et al.[17]	Heuristic Miner			
Trcka and Pechenizkiy[7]		Conformance checker		
Southavilay et al.[39]	Heuristic Miner		X	
Trcka et al.[1]	Fuzzy Miner	LTL—Conformance checker	X	
Poncin et al.[49]	Fuzzy Miner		X	
Ayutaya et al.[47]	Heuristic Miner			
Anuwatvisit et al.[48]		Conformance checker		
Schoor and Bannert[37]	Fuzzy Miner			
van der Aalst et al.[12]	Fuzzy Miner	Conformance checker	X	
Reimann et al.[8]	Fuzzy Miner			
Barreiros et al.[24]	Genetic Algorithm			
Bannert et al.[18]	Fuzzy Miner	LTL—Conformance checker		
Cairns et al.[40]	Heuristic Miner	LTL		
Cairns et al.[32]	Fuzzy Miner			X
Bogarin et al.[26]	Heuristic Miner			
Cairns et al.[31]		LTL—Conformance checker	X	
Cairns et al.[13]	Fuzzy Miner	LTL—Conformance checker	X	X
Mukala et al.[34]	Fuzzy Miner	Conformance checker	X	
Vahdat et al.[42]	Fuzzy Miner			
Ariouat et al.[43]	Heuristic Miner			
Okoye et al.[55]	Fuzzy Miner			
Sedrakyan et al.[54]	Fuzzy Miner		X	
Vidal et al.[21]	Genetic Algorithm			

support for dynamic role assignment.[36] Other authors like Schoor and Bannert[37] have explored sequences of social regulatory processes during a CSCL task and their relationship to group performance. This study used PM to identify process patterns for high and low group-performance dyads. In the most recent research in this domain, Porouhan and Premchaiswadi[38] apply several PM techniques such as social network mining, basic performance analysis, role hierarchy mining, and dotted chart analysis with the aim of increasing the instructor's awareness/knowledge about the collaborative dynamics in each group. A particular application of EPM to this domain is Collaborative Writing (CW). CW is widely used in education environments where students often use computers to take notes during lectures and write essays for their assignments. Thanks to the availability of the Internet, students can also write collaboratively by sharing their documents in a number of ways. PM has been used in Southavilay et al.[39] to analyze students' writing processes and how these processes correlate to the quality and semantic features of the final product. They used documents collected from different groups of undergraduate students writing collaboratively in order to evaluate the proposed heuristics and illustrate the applicability of PM techniques to analyzing the writing process.

Professional Training

Education and training centers have made their professional training courses more agile in order to respond to the changing needs of the job market and meet time-to-skill requirements.[32]

PM has been used in different types of professional training courses. Cairns et al.[32] showed how PM can be used to monitor and improve educational processes in the field of professional training. Their research aims to develop generic methods which could be applied to general education issues and more specific applications concerning professional

TABLE 7 | Targets in EPM Educational Application Domains

Application	Work/Paper	Target
MOOCs, LMS, and hypermedia environments	Mukala et al.[34]	To detect learning difficulties
	Mukala et al.[33]	To generate recommendations or advice for students.
	Bogarin et al.[26]	To gain a better understanding of the underlying educational process
	Vidal et al.[21]	To improve management of learning objects
	Bannert et al.[18]	To detect learning difficulties and discover sequential patterns
	Reimann et al.[8]	To discover sequential patterns
	Trcka et al.[1]	To discover learning flows
	Emond and Buffett.[35]	
Computer-supported collaborative learning	Reimann et al.[17]	To discover learning flows and provide feedback
	Bergenthum et al.[9]	To discover learning flows
	Schoor and Bannert[37]	To discover sequential patterns
	Porouhan and Premchaiswadi[38]	To gain a better understanding of the underlying educational process
	Southavilay et al.[39]	To gain a better understanding of the underlying educational process and detect learning difficulties
Professional training	Cairns et al.[32]	To analyze social networks
	Cairns et al.[31]	To discover learning flows
	Doleck et al. 2016[41]	To gain a better understanding of the underlying educational process and detect learning difficulties
	Vahdat et al.[42]	
	Ariouat et al., 2016[43]	To gain a better understanding of the underlying educational process
Curriculum mining	Trcka and Pechenizkiy[7]	To gain a better understanding of the underlying educational process
	Wang and Zaïane[44]	To gain a better understanding of the underlying educational process and generate recommendations or advice for students.
	Schulte et al.[45]	To generate recommendations or advice for students.
Computer-based assessment	Pechenizkiy et al.[20]	To provide feedback
	Tóth et al.[46]	To detect learning difficulties
Student registration	Anuwatvisit et al.[48]	To gain a better understanding of the underlying educational process
	Ayutaya et al.[47]	To discover learning flows
Software repositories	Poncin et al.[50]	To gain a better understanding of the software development process
	Poncin et al.[49]	
Structured inquiry cycle	Howard et al.[51]	To detect learning difficulties
	Jeong et al.[52]	
3D educational virtual worlds	Fernández-Gallego et al.[53]	To discover learning flows

training or e-learning fields for the extraction, analysis, enhancement and personalization of educational processes. In similar research, Cairns et al.[31] analyzed training processes and their conformance with established curriculum constraints, educators' hypotheses and prerequisites, and to enhance training process models with performance indicators such as execution time, bottlenecks and decision points. Doleck et al.[41] explored knowledge-based discovery approaches to understand learner behaviors in a medical computer-based learning environment using PM techniques in order to provide a more coherent picture about clinical diagnosis reasoning. Vahdat et al.[42] carried out their study in a simulation environment for learning digital electronics. They exploited PM methods to investigate and compare the learning processes of professionals. In order to do that, they measured the understandability of their process models through a complexity metric. Additionally, Ariouat et al.[43] tried to identify the best training paths in real-world professional training databases from a global consulting company.

Curriculum Mining

A curriculum is partially designed by an educational institution in order to accomplish certain goals. Curricula normally suggest that students can follow differing paths from start to end due to a liberal approach in selecting courses.[44]

A domain-driven EPM approach was proposed by Trcka and Pechenizkiy[7] in curriculum mining. They proposed a framework which assumes that a set of pattern templates can be predefined to focus the mining in a desired way and make it more effective and efficient. The framework is aimed at helping educators analyze educational processes based on formal modeling. In other related research, Wang and Zaïane[44] discovered a curriculum process model of students taking courses and compared the paths that successful and less successful students tended to take, highlighting discrepancies between them. In other work Schulte et al.[45] presented research into EPM and student data analytics in a whole university scale approach with the aim of providing insight into questions raised by degree pathways. Their goal was to uncover statistically significant and meaningful patterns in students' course pathway choices, and to provide student support units, degree and course coordinators with longitudinal indicators that could be used to inform students.

Computer-based Assessment

Computer-based assessment (CBA) is, in essence, the practice of giving quizzes and tests on the computer instead of using traditional pencil and paper formats. Computer based assessment is widely used in many different VLEs.

PM has been used to exclusively analyze assessment data coming from two online multiple choice test studies showing the utility of process discovery, conformance checking and performance analysis techniques.[20] In a similar context, Tóth et al.[46] described how to extract process-related information from event logs, and how to use these data in problem-solving assessments and describe methods which help discover novel information based on individual problem-solving behavior.

Student Registration

Student registration deals with all the requirements and steps of the academic or course registration process. It is critical to check on automated management system processes in the educational domain in order to produce expected results in terms of quality and timely student registration processes.[47]

In this context, Ayutaya et al.[47] used the Heuristic Mining technique to gain insight into student registration processes at a Thai university. The most important characteristic of Heuristics Miner is its robustness against noise and exceptions. Because Heuristics Miner is based on the frequency of patterns it is possible to focus on the main behavior in the event log and that makes it especially appropriate for unstructured educational processes. Anuwatvisit et al.[48] used Conformance checker in order to detect discrepancies between the flows prescribed in a student registration model and the actual process instances. In addition, they extended the models with performance characteristics and business rules.

Software Repositories

Developers and development teams are involved in software development processes, often from different locations. In these projects, different kinds of software repositories such as source-code management systems, document repositories, mail archives, bug trackers and version control systems are used to support communication and coordination.

PM has also been applied to mining software repositories. Poncin et al.[49] identified the challenges that need to be addressed to enable this application, discussing how they can be addressed and presented through FRASR (Framework for Analyzing Software Repositories). PM has been also applied in Poncin et al.[50] to analyze data from multiple software repositories. The preprocessing step extracts information from software repositories, which have different structures, and combines information into an event log, while the analysis step is aimed at discovering the process structure, reflected in the log, and analyzing whether it is correct or visualizing it.

Structured Inquiry Cycle

A structured inquiry cycle is a kind of adaptation strategy that combines explicit structuring and scaffolding with increased learner control to create a freer, more personalized learning experience due to high potential variation in prior knowledge, metacognitive skills, and motivation within learner populations.[51]

A PM approach using a structured inquiry cycle has been applied in a corpus of online modules for adult informal learners.[51] Informal learning situations often exhibit high variability within the learner population, especially when learning experiences and environments offer broad availability. However, this freedom of navigation sometimes has negative effects on learning experiences, particularly when prior domain knowledge or learning skills are weak. The authors demonstrated that Petri Net process models which aid collaborative planning and reviewing results, where unshared, informal understanding can be a hindrance.

In a similar context, Jeong et al.[52] used a hidden Markov model approach for exploratory sequence analysis by applying the methodology to studying student learning behaviors in a new domain that promotes an inquiry cycle.

3D Educational Virtual Worlds

3D Educational Virtual Worlds are environments that encourage interaction between students and teachers. These environments encourage students (as avatars) to perform learning activities that were not initially scheduled by the teachers.

PM has been used in order to find out what is happening in 3D student learning processes. With that aim, Fernández-Gallego et al.[53] presented a learning analytics framework for 3D educational virtual worlds that focuses on discovering learning flows and checking conformance through PM techniques because, in this specific domain, the interactions among students are continuous and there is a lot of noise, in other words, a high number of activities that are not significant from a pedagogical point of view.

Finally, Table 7 shows a summary of all the previously described EPM research and its target grouped by application domain. On the one hand, we can see that currently, the most active domains are: Research into MOOC, LMS and Hypermedia environments, discovering in CSCL, and discovering in Professional Training. On the other hand, the results of EPM can be used to gain a better understanding of the underlying educational processes, to provide feedback to students, teachers and researchers, to detect learning difficulties and to help students with specific learning disabilities, to improve management of learning objects, to generate advice for students, among many other uses. Table 7 shows the general target of every study indicating that the most frequent targets in current EPM research are focused on gaining a better understanding of underlying educational processes, detecting learning difficulties and discovering learning flows.

CONCLUSION AND FUTURE WORKS

This paper presents a comprehensive introduction to EPM, one of the most promising EDM techniques. EPM is young, emerging field which is closely related to other research areas such as IM, SPM and G-EDM. The present work describes the EPM framework which is an adaptation of the generic PM framework and addresses the different hurdles that may arise when handling event logs, the most commonly used techniques, and most often used tools. Additionally, it provides an overview of the main research to date that aims to be a guide to the available research in this regard.

EPM allows a better understanding of the underlying educational process from raw event data but as an emerging area, it faces many challenges and has many prospects for the future. We would like to outline some important issues that will be especially challenging in EPM's near future, such as:

- **Development of more specific EPM tools** in order to bring PM closer to the domain experts (i.e., educational specialists and researchers, who do not necessarily have all the technical background), helping them to analyze educational processes. SoftLearn[24] is currently the only specific EPM tool. Additionally, PHIDIAS is a project developed by ALTRAN researchers[32] with the aim of developing an interactive platform tailored for educational process reconstruction and analysis. This platform will allow different education centers and institutions to load their data and provide access to advanced DM and PM services.
- **Using semantics to improve EPM.** Semantic concepts can be layered on top of existing learner information assets to provide a more conceptual analysis of real time processes capable of providing real world answers that are closer to human understanding. Along these lines, Cairns et al.[32] proposed how linking labels in event logs to their underlying semantics can bring educational process discovery to the conceptual level. In this way, more accurate and compact educational processes can be mined and analyzed at different levels of abstraction. Sedrakyan et al.[54] focused on the modeling activities that can potentially affect the semantic quality of a conceptual model process. They constructed a semantically correct conceptual model that reflects the structural and

dynamic view of a given domain description. Finally, Okoye et al.[55] proposed semantic PM to enrich streams of event data logs from a learning process using semantic descriptions that references concepts in an ontology that is specifically designed for representing learning processes.

- **Using recommendation in EPM.** The information discovered in PMs has to be, not only comprehensible, but also useful for the end-users' decision-making. For example, instead of showing the overall process model obtained, it is better to abstract its representation using, for example, the normal academic notation understandable to end-users[8,18] or in the form of a list of suggestions, recommendations, and conclusions about the results.[13] So, in addition of the three main types of EPM, some authors,[14,32] consider recommendation as another emerging type. Recommendation can be seen an extension of the enhancement type instead of a new type of EPM. Chen et al.[56] used PM to extract and reorganize the learning process in order to be able to recommend a learning unit for the target user to learn as the next step. Wang and Zaïane,[44] used PM to adjust the requirements for the curriculum and to recommend courses to students based on expected outcome. Cairns et al.[31] show recommendations of the best course units or learning paths to students (depending on their profiles, preferences or target skills) and the on-line detection of unmet prerequisites.
- **The application of EPM to other emergent educational domains** such as games, mobile, and ubiquitous learning environments. The play traces resulting from the learner's activity in learning games are hard for teachers to analyze and interpret. EPM techniques have started to be applied in Learning Games applications[57] in order to analyze of student behaviors. Digital mobile devices such as tablets, PDAs, and smart phones are also being used increasingly often for educational purposes. Process Mining and Learner Behavior Analytics[38] have been used in a Collaborative and Web-Based MultiTabletop Environment.
- **Make more free EPM datasets public** in order to test theoretical models and assumptions by methods of PM. Free EPM datasets could be very useful for testing some *ad-hoc* models. However, not all the current EPM datasets used in research are available to download. We believe it is essential that EPM datasets are freely available in order for EPM research to grow much more quickly. In fact, we have only found one free specific EPM dataset (https://archive.ics.uci.edu/ml/datasets/Educational+Process+Mining+(EPM)%3A+A+Learning+Analytics+Data+Set). Therefore, one of the most important next steps is to promote the sharing of online data-collections that can be analyzed from multiple perspectives, with various methods and tools.

ACKNOWLEDGMENTS

This research is supported by project TIN2014-55252-P from the Spanish Ministry of Science and Technology and the Department of Science and Innovation under the National Program for Research, Development, and Innovation: EDU2014-57571-P.We have also received funds from the European Union, through European Regional Development Funds (ERDF); and the Principality of Asturias, through its Science, Technology and Innovation Plan (grant GRUPIN14-053).

REFERENCES

1. Trcka N, Pechenizkiy M, Van der Aalst WMP. *Process Mining from Educational Data* (Chapter 9); 2011.
2. Van der Aalst W, Weijters T, Maruster L. Workflow mining: discovering process models from event logs. *IEEE Trans Knowl Data Eng* 2004, 16:1128–1142.
3. Romero C, Ventura S. Educational data science in massive open online courses. *WIREs Data Mining Knowl Discov* 2017, 7:1–12.
4. Weijters AJMM, van Der Aalst WM, De Medeiros AA. Process mining with the heuristics miner-algorithm. In: *Technische Universiteit Eindhoven, Tech. Rep. WP*. Vol *166*, 2006, 1–34.
5. Romero C, Cerezo R, Bogarín A, Sánchez-Santillán M. Educational process mining: a tutorial and case study using moodle data sets. In: *Data Mining and Learning Analytics: Applications in Educational Research*. Hoboken, NJ: John Wiley & Sons; 2016, 1–28.
6. Mans RS, van der Aalst W, Vanwersch RJ. *Process Mining* in Healthcare: Evaluating and Exploiting Operational Healthcare Processes. Berlin, Germany: Springer; 2015, 17–26.

7. Trcka N, Pechenizkiy M. From local patterns to global models: towards domain driven educational process mining. In: *Ninth International Conference on Intelligent Systems Design and Applications*, IEEE, Pisa, Italy, 2009, 1114–1119.
8. Reimann P, Markauskaite L, Bannert M. E-research and learning theory: what do sequence and process mining methods contribute? *Br J Educ Technol* 2014, 45:528–540.
9. Bergenthum R, Desel J, Harrer A, Mauser S. Modeling and mining of learnflows. *Trans Petri Nets Other Models Concurrency* 2012, 5:22–50.
10. Perez-Rodriguez R, Caeiro-Rodriguez M, Anido-Rifon L. Enabling process-based collaboration in Moodle by using aspectual services. In: *Ninth IEEE International Conference on Advanced Learning Technologies*, IEEE, Riga, Latvia, 2009, 301–302.
11. Romero C, Ventura S. Data mining in education. *WIREs Data Mining Knowl Discov* 2013, 3:12–27.
12. Van der Aalst WM, Guo S, Gorissen P. Comparative process mining in education: An approach based on process cubes. In: *International Symposium on Data-Driven Process Discovery and Analysis*. Springer Berlin Heidelber, Riva del Garda, Italy; 2013, 110–134.
13. Cairns AH, Gueni B, Assu J, Joubert C, Khelifa N. Process mining in the education domain. *Int J Adv Intell Syst* 2015, 8:219–232.
14. Khodabandelou G, Hug C, Deneckere R, Salinesi C. Process mining versus intention mining. In: *Enterprise, Business-Process and Information Systems Modeling*. Springer Berlin Heidelberg,Valencia, Spain; 2013, 466–480.
15. Agrawal R, Srikant R. Mining sequential patterns. In: *Data Engineering Proceedings of the Eleventh International Conference*. IEEE, Taipei, Taiwan; 1995, 3–14.
16. Nesbit JC, Zhou M, Xu Y, Winne P. Advancing log analysis of student interactions with cognitive tools. In: *12th Biennial Conference of the European Association for Research on Learning and Insruction (EARLI)*. Budapest, Hungary; 2007, 2–20.
17. Reimann P, Frerejean J, Thompson K. 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, Vol. 1*. Rhodes, Greece; 2009, 98–107.
18. Bannert M, Reimann P, Sonnenberg C. Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacogn Learn* 2014, 9:161–185.
19. Washio T, Motoda H. State of the art of graph-based data mining. *Acm Sigkdd Explor Newslett* 2003, 5:59–68.
20. Pechenizkiy M, Trcka N, Vasilyeva E, van Aalst W, De Bra P. Process mining online assessment data. In: *Educational Data Mining*. Córdoba, Spain; 2009, 279–288.
21. Vidal JC, Vázquez-Barreiros B, Lama M, Mucientes M. Recompiling learning processes from event logs. *Knowledge-Based Syst* 2016, 100:160–174.
22. Van der Aalst WM. *Process Mining: Data Science in Action*. Berlin, Germany:Springer; 2016.
23. Schoonenboom J, Levene M, Heller J, Keenoy K, Turcansyi M. *Trails in Education: Technologies that Support Navigational Learning*. Rotterdam, The Netherlands: Sense Publishers; 2007.
24. Barreiros BV, Lama M, Mucientes M, Vidal JC. Softlearn: a process mining platform for the discovery of learning paths. In: *14th International Conference on Advanced Learning Technologies*. IEEE, Athens, Greece; 2014, 373–375.
25. Mekhala A. Review paper on process mining. *Int J Eng Res Technol* 2015, 1:11–17.
26. Bogarín A, Romero C, Cerezo R, Sánchez-Santillán M. Clustering for improving educational process mining. In: *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. ACM, Indianapolis, USA; 2014, 11–15.
27. Günther CW, Van Der Aalst WM. Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In: *International Conference on Business Process Management*. Springer Berlin Heidelberg; Brisbane, Australia; 2007, 328–343.
28. Vidal JC, Lama M, Bugarín A. Petri net-based engine for adaptive learning. *Expert Syst Appl* 2012, 39:12799–12813.
29. Van Dongen BF, de Medeiros AKA, Verbeek HMW, Weijters AJMM, Van Der Aalst WM. The ProM framework: a new era in process mining tool support. In: *International Conference on Application and Theory of Petri Nets*. Springer Berlin Heidelberg; Miami, USA; 2005, 444–454.
30. Rozinat A, van der Aalst WM. Conformance testing: Measuring the fit and appropriateness of event logs and process models. In: *International Conference on Business Process Management*. Springer Berlin Heidelberg; Nancy, France; 2005, 163–176.
31. Cairns AH, Gueni B, Assu J, Joubert C, Khelifa N. Analyzing and improving educational process models using process mining techniques. In: *The Fifth International Conference on Advances in Information Mining and Management*. Brussels, Belgium; 2015, 17–22.
32. Cairns AH, Gueni B, Fhima M, Cairns A, David S, Khelifa N. Towards custom-designed professional training contents and curriculums through educational process mining. In: *The Fourth International Conference on Advances in Information Mining and Management*. Paris, France; 2014, 53–58.
33. Mukala P, Buijs J, van der Aalst WMP. Uncovering learning patterns in a MOOC through conformance alignments. In: Tech. rep., Eindhoven University of Technology, BPM Center Report BPM; 2015.

34. Mukala P, Buijs J, Leemans M, van der Aalst W. Learning analytics on coursera event data: a process mining approach. In: *Proceedings of the 5th International Symposium on Data-driven Process Discovery and Analysis*. Vienna, Austria; 2015, 18–32.
35. Emond B, Buffett S. Analyzing student inquiry data using process discovery and sequence classification. In: *Proceedings of the 8th International Conference on Educational Data Mining*. Madrid, Spain; 2015, 412–415.
36. Bergenthum R, Desel J, Harrer A, Mauser S. Learnflow mining. *DeLFI* 2008, 132:269–280.
37. Schoor C, Bannert M. Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Comput Hum Behav* 2012, 28:1321–1331.
38. Porouhan P, Premchaiswadi W. Process mining and learners' behavior analytics in a collaborative and web-based multi-tabletop environment. *Int J Online Pedagogy Course Design* 2017, 7:29–53.
39. Southavilay V, Yacef K, Callvo RA. Process mining to support students' collaborative writing. In: *Proceedings of the 3th International Conference on Educational Data Mining*, Pittsburgh, USA, 2010:257–266.
40. Cairns AH, Ondo JA, Gueni B, Fhima M, Schwarcfeld M, Joubert C, Khelifa N. Using semantic lifting for improving educational process models discovery and analysis. In: *SIMPDA*. Milan, Italy; 2014, 150–161.
41. Doleck T, Jarrell A, Poitras EG, Chaouachi M, Lajoie SP. Examining diagnosis paths: a process mining approach. In: *Computational Intelligence & Communication Technology (CICT), Second International Conference*. IEEE, Ghaziabad, India, 2016, 663–667.
42. Vahdat M, Oneto L, Anguita D, Funk M, Rauterberg M. A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In: *Design for Teaching and Learning in a Networked World*. Toledo, Spain: Springer International Publishing; 2015, 352–366.
43. Ariouat H, Cairns AH, Barkaoui K, Akoka J, Khelifa N. A two-step clustering approach for improving educational process model discovery. In: *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 25th International Conference*. IEEE; Paris, France; 2016, 38–43.
44. Wang R, Zaïane OR. Discovering process in curriculum data to provide recommendation. In: *Proceedings of the 5th International Conference on Educational Data Mining,* Madrid, Spain, 2015, 580–581.
45. Schulte J, Fernandez de Mendonca P, Martinez-Maldonado R, Buckingham Shum S. Large scale predictive process mining and analytics of university degree course data. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM; Vancouver, Canada; 2017, 538–539.
46. Tóth K, Rölke H, Goldhammer F, Barkow I. Educational process mining: new possibilities for understanding students' problem-solving skills. In: *Educational Research and Innovation*. Paris, France: OECD Publishing; 2017, 193–209.
47. Ayutaya NSN, Palungsuntikul P, Premchaiswadi W. Heuristic mining: Adaptive process simplification in education. In: *ICT and Knowledge Engineering (ICT & Knowledge Engineering), 10th International Conference*. IEEE; Bangkok, Thailand; 2012, 221–227.
48. Anuwatvisit S, Tungkasthan A, Premchaiswadi W. Bottleneck mining and Petri net simulation in education situations. In: *ICT and Knowledge Engineering (ICT & Knowledge Engineering), 10th International Conference*. IEEE; Bangkok, Thailand; 2012, 244–251.
49. Poncin W, Serebrenik A, van den Brand M. Process mining software repositories. In: *15th European Conference on Software Maintenance and Reengineering (CSMR)*. IEEE; Burlington, USA; 2011, 5–14.
50. Poncin W, Serebrenik A, van den Brand M. Mining student capstone projects with FRASR and ProM. In: *Proceedings of the ACM International Conference Companion on Object Oriented Programming Systems Languages and Applications Companion*. Portland, USA; 2011, 87–96.
51. Howard L, Johnson J, Neitzel C. Examining learner control in a structured inquiry cycle using process mining. In: *Proceedings of the 3th International Conference on Educational Data Mining*, Pittsburgh, USA, 2010:71–80.
52. Jeong H, Biswas G, Johnson J, Howard L. Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models. In: *Proceedings of the 3th International Conference on Educational Data Mining*, Pittsburgh, USA, 2010:81–90.
53. Fernández-Gallego B, Lama M, Vidal JC, Mucientes M. Learning analytics framework for educational virtual worlds. *Proc Comput Sci* 2013, 25:443–447.
54. Sedrakyan G, De Weerdt J, Snoeck M. Process-mining enabled feedback: "tell me what I did wrong" vs. "tell me how to do it right". *Comput Hum Behav* 2016, 57:352–376.
55. Okoye K, Tawil ARH, Naeem U, Lamine E. Discovery and enhancement of learning model analysis through semantic process mining. *Int J Comput Inform Syst Industrial Manage Appl* 2016, 8:93–114.
56. Chen J, Zhang Y, Sun J, Chen Y, Lin F, Jin Q. Personalized micro-learning support based on process mining. In: *7th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE; Huangshan, Anhui, China; 2015, 511–515.
57. Vermeulen M, Mandran N, Labat JM. Chronicle of a scenario graph: from expected to observed learning path. In: *11th European Conference on Technology Enhanced Learning*. Springer Berlin Heidelberg, Lyon, France; 2016, 321–330.