2014 IEEE 14th International Conference on Advanced Learning Technologies

SoftLearn: A Process Mining Platform for the Discovery of Learning Paths

Borja Vázquez-Barreiros, Manuel Lama, Manuel Mucientes, Juan C. Vidal
Center for Research in Information Technologies (CiTIUS)
University of Santiago de Compostela, Spain
email: borja.vazquez@usc.es, manuel.lama@usc.es, manuel.mucientes@usc.es, juan.vidal@usc.es

Abstract—**One of the most challenging issues in learning analytics is the development of techniques and tools that facilitate the evaluation of the learning activities carried out by learners. In this paper, we faced this issue through a process mining-based platform, called SoftLearn, that is able to discover complete, precise and simple learning paths from event logs. This platform has a graphical interface that allows teachers to better understand the real learning paths undertaken by learners.**

Keywords-learning analytics; process mining; learning path discovery

I. INTRODUCTION

In the last two decades a great effort has been made in developing virtual learning environments (VLEs) that allow to manage the learning activities and contents of a course. However, little attention has been paid to develop techniques for understanding the learners' behavior when they carry out the sequence of learning activities, i.e., the learning path, which was planned for achieving the pedagogical objectives of a course [1]. Thus key questions related to this behavior remain still open, e.g., have learners followed the learning paths designed by the teacher? In this context, the issue is to determine whether learners have undertaken additional learning activities, such as looking for new learning contents, in order to better understand the learner's behavior. These activities should be highlighted to enable teachers to improve the *learning paths* as well as the *evaluation process* [2].

To discover the learners' behavior, authors have applied two different kinds of techniques: *sequential pattern mining* (SPM) [3] and *process mining* (PM) [4]. On one hand, SPM techniques are mainly oriented to discover *simple* behavioral patterns that could occur when learners undertake the learning activities of a course, such as in self-regulated learning [5] or collaborative learning [6]. Therefore, SPM is not appropriate to discover learning paths that describe the *whole learning process* of a course [7]. On the other hand, PM techniques have been applied to automatically discover the real workflow of learning activities that learners have undertaken. PM techniques achieve this objective by analyzing the events generated as consequence of the learner's activity in the VLE. Thus some authors have applied PM to discover the learning processes in self-regulated learning [7], collaborative writing [8] or multiple-choice questions tests [9]. These works have two main drawbacks. The first one is that they do not assure that all the activities undertaken by learners are included in the learning path. The second drawback is that these works do not provide a graphical tool that allow teachers to visualize the real learning path of a course as well as to access easily to the learning content generated in the VLE.

In this paper, we present a process mining-based platform, known as SoftLearn, which uses a genetic algorithm to discover *complete* learning paths, guaranteeing that there are no missing activities and therefore enabling teachers to use this platform for *evaluating* the learning activities. Furthermore, the SoftLearn platform has an intuitive graphical interface which has been specifically developed to visualize both the learning paths discovered by the genetic algorithm and the data generated during the learning activities.

II. SOFTLEARN PLATFORM

The software architecture of the SoftLearn platform is depicted in Fig. 1. The first component of this architecture is the virtual learning environment where learners and teachers carry out the learning and support activities planned for a course, access to the learning contents, and execute the services needed to undertake those activities. The only one restriction required for the VLE is that it must register the user activity that takes place in the virtual environment. This restriction is not hard, since most of the current VLEs include some kind of monitoring of the learner behavior. Hereinafter, the remaining SoftLearn components are described.

A. Event Log System

When a learner undertakes a learning activity, such as creating a blog input or answering a test, the VLE stores in a database the information generated as result of performing this learning activity. However, in order to be able to obtain the graph that represents the learning path of the course, it is necessary to translate this information into the input format of the process discovery algorithm. This translation is carried out by VLE-specific *adapters* that provide the independence of SoftLearn with the particular VLE in which the learning-teaching process takes place.

B. Process Mining Algorithm

The process mining algorithm (PMA) is the core component of the SoftLearn platform, since its aim is to *discover* the workflow that represents the learning path followed by the learners during a course. To achieve this objective the PMA only needs to process the metadata information that is available in the event log, which is independent of the kind of learning activities undertaken by learners.

From the perspective of process mining, the quality of a discovery algorithm is measured taking into account the following metrics:

978-1-4799-4038-7/14 $31.00 © 2014 IEEE
DOI 10.1109/ICALT.2014.111

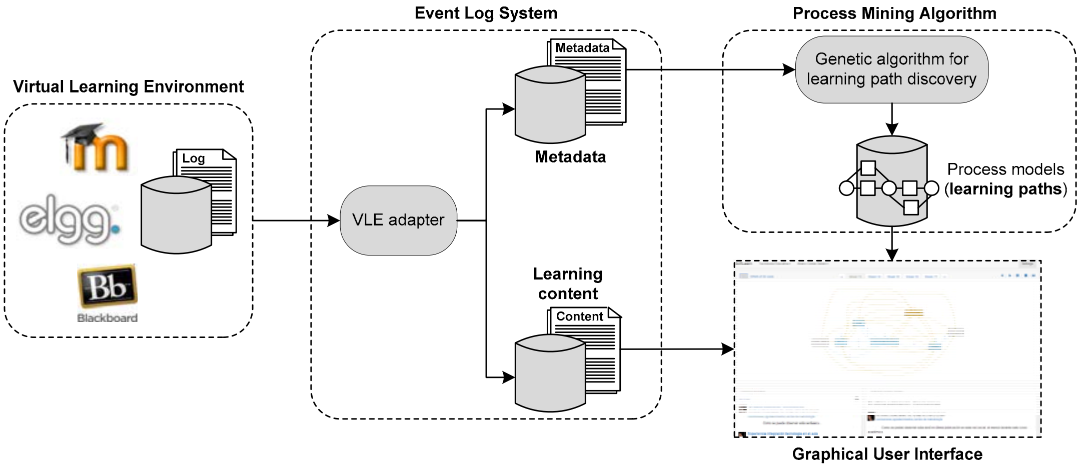

Figure 1. SoftLearn software architecture.

- *Completeness*: a discovered learning path is considered as *complete* when it can reproduce all the events contained in the log database. In order to guarantee feasible and correct evaluations of the learning paths, teachers need to access to all the activities performed by learners. Therefore, completeness is a *hard requirement* for the PMA.
- *Precision*: a discovered learning path is considered as precise when it cannot reproduce events that are not available in the log database. From the point of view of the learner's evaluation, this kind of learning paths are desirable, but it is not a requirement as hard as completeness: teachers will not visualize the extra behavior because it has not been undertaken by any learner.
- *Simplicity*: a discovered learning path is considered simple when it presents the minimal structure that reflects the behavior stored in the log database. A desirable requirement for the PMA is to obtain simple representations of the discovered learning paths to improve the understanding of the learner's behavior.

Taking these measures into account, we have developed a genetic algorithm, known as ProDiGen, for process mining in VLE that obtains *complete* learning paths with *very high* values for precision and simplicity.

C. SoftLearn Graphical Interface

The graphical user interface of SoftLearn is the component (*http://tec.citius.usc.es/SoftLearn*) that enables teachers to understand the learner's behavior through the visualization of learning paths followed by the learners during a course. Fig. 2 shows a screenshot of the graphical interface for a course where the learning activities are undertaken in a social network supported by ELGG.

The visualization of learning paths with the aim of facilitating the course evaluation must consider two main issues: how to *represent learning paths* and how to *present the data* about the learning activities, including the contents generated by learners as part of these activities.

1) Graph representation of learning paths

As there is no standard language to visualize learning designs, we decided to show learning paths through D/F-graphs [4]. This kind of graphs represents dependencies between learning activities with arcs, meaning that the source activity of an arc is carried out *before* the target activity of that arc. As D/F-graphs do not contain any control structure, they are easy to understand.

When teachers select a temporal period, such as a week, a month, two months or the whole course, the learning path that describes the behavior for *all the learners* is presented. Then, if the teacher selects a particular learner (Fig. 2.c), the learning activities performed by him are highlighted (blue-filled rectangles in Fig. 2.a). Furthermore, in order to facilitate the learning path visualization, the graphical interface incorporates a *graph player* that allows the execution of the graph in several ways (Fig. 2.b):

- *Step by step mode* enables users to navigate through the dependency graph by following the execution of each learning activity performed by the learner. This mode is specially designed for the visualization of the learning content associated to the activity which has been activated in the current step.
- *Play mode*, where the learning path is executed from the beginning to the final activity undertaken by the learner. Users can set up the speed at which the activities are highlighted.
- *Player actions* that enable teachers to *stop*, *resume*, and *reset* the execution of the learning path executed at each moment. These actions can be performed in both *step by step* and *play* modes.

2) Visualization of the learning activity data

The dependency graph presents the learning path related to the learner's behavior, but it does not show any data about the learning activities of that graph. These data are needed to make an effective learner evaluation. The graphical interface displays (i) *metadata*, which describe the features of the learning activities, including their timestamp, snippet, creator name, etc (Fig. 2.d); and (ii) *learning content*, which is related with the learning activity description and with the resources content generated by the learner in the context of the learning activities (Fig. 2.e).

III. CONCLUSIONS

The core component of the SoftLearn platform is the genetic algorithm that discovers *complete* learning paths, i.e., learning paths that can process every event stored in the log register. Therefore, evaluations of the learner's behavior are reliable, since there are no missing activities. This is the basis for developing a graphical interface that teachers can use to visualize learning paths as activity graphs and to access to the relevant data generated in the learning activities.

As future work new visualization functionalities will be integrated within the graphical interface, such as comparing learning paths among different groups of learners or browsing activity graphs based on calendar views.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Economy and Competitiveness under the project TIN2011-22935 and by the European Regional Development Fund (ERDF/FEDER) under the project CN2012/151 of the Galician Ministry of Education.

REFERENCES

[1] Romero, and S. Ventura, "Data mining in education," Wiley Interdisc. Rew.: Data Min. Knowl. Disc., vol. 3, 2013, 12–27.

[2] C. Romero, S. Ventura, and E. García, "Data Mining in Course Management Systems: Moodle Case Study and Tutorial, " Computers & Education, vol. 51, 2008, pp. 368–384.

[3] R. Agrawal, and R. Srikant, "Mining sequential patterns," Proc. 11th IEEE Int. Conf. Data Engineering (ICDE 95), 1995, pp. 3–14.

[4] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer-Verlag, 2011.

[5] F. Bouchet, J. S. Kinnebrew, G. Biswas, and R. Azevedo, "Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System fostering Self-Regulated Learning," Proc. 5th Int. Conf. Educational Data Mining (EDM 12), 2012, pp. 65–72.

[6] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaïane, "Clustering and Sequential Pattern Mining of Online Collaborative Learning Data," IEEE Trans. Knowl. Data Eng., vol. 21, 2009, pp. 759–772.

[7] M. Bannert, P. Reimann, and C. Sonnenberg, "Process Mining Techniques for Analysing Patterns and Strategies in Students' Self-Regulated Learning," Metacognition and Learning, vol 8, 2013.

[8] R. A. Calvo, S. T. O'Rourke, J. Jones, K. Yacef, and P. Reimann, "Collaborative Writing Support Tools on the Cloud," IEEE Trans. Learn. Technol., vol. 4, 2011, 88–97.

[9] M. Pechenizkiy, N. Trcka, E. Vasilyeva, W. M. P. van der Aalst, and P. de Bra, "Process Mining Online Assessment Data," Proc. 2nd Int. Conf. Educational Data Mining (EDM 09), 2009, pp. 279–288.

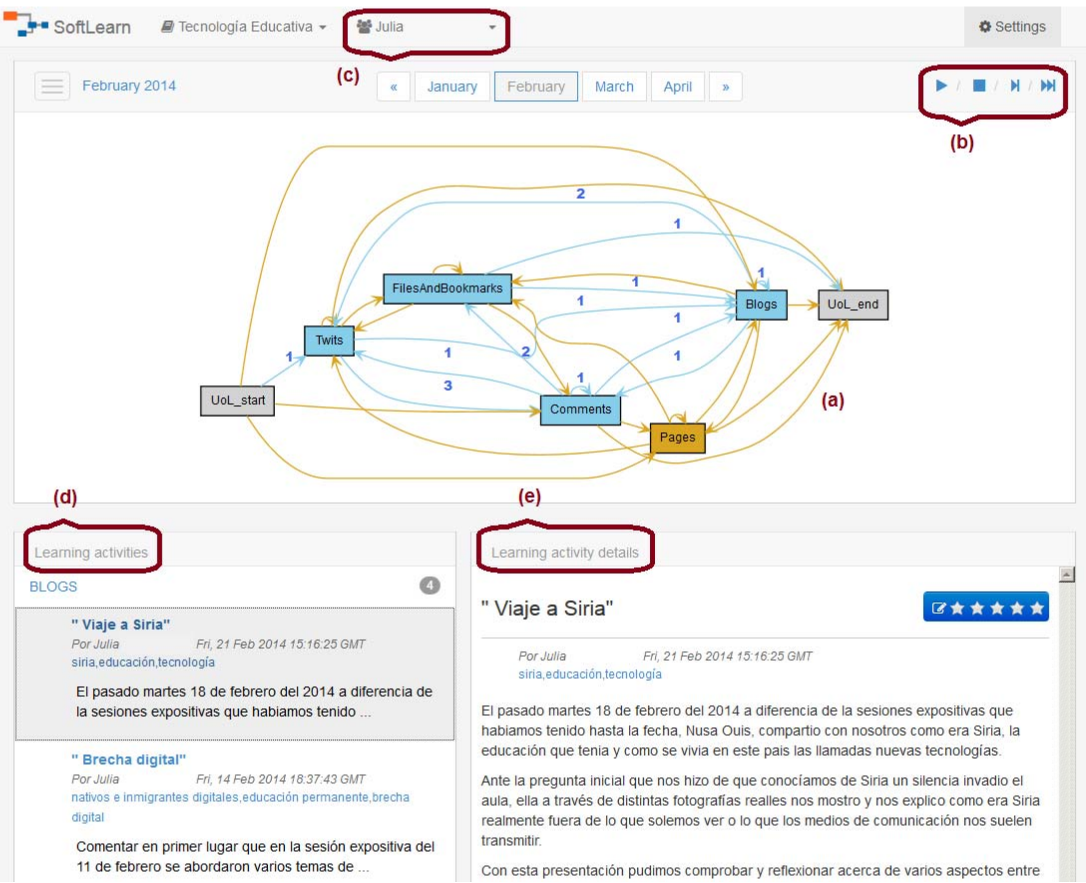

Figure 2. SoftLearn graphical interface.