DATA EXPLOITATION IN INTEGRATED DEVELOPMENT ENVIRONMENTS USED IN COMPUTER SCIENCE TEACHING

N. El Bahri¹, Z. Itahriouan², A. Abtoy¹, S. Brahim Belhaouari³

¹Laboratoire Informatique, Recherche Opérationnelle et Statistiques Appliquées, Abdelmalek Essadi University (MOROCCO)
²Laboratoire Systèmes et Environnements Durables, Private University of Fez (MOROCCO)
³Division of Information and Computing Technology, College of Science & Engineering Hamad Bin Khalifa University (QATAR)

Abstract

In web 2.0 era, the use of technology to learn has benefited from all its influence giving therefore birth of the e-learning 2.0 concept. As a result, teaching and learning are eventually taking place in social and interactive environments giving a virtual presence of the actors (teacher, student, etc.) as well as their interactions. These latter are often represented as data, which open up promising opportunities for their exploitation to improve teaching and learning operations.

Nowadays, Learning Management Systems, Massive Open Online Courses and Tutoring Systems are the most exploited environment to extract useful data. Several researches are conducted to exploit data and interactions in measuring learner engagement, predicting analysis, early detection and many other applications.

In another context and evermore in E-learning 2.0 field, our team has developed a collaborative social web IDE (Integrated Development Environment) to be used in practical computer science teaching. This environment virtualizes the sequence of practical work. It also enables the collaboration between student and teacher and students themselves. The use of our IDE in practical programming work has demonstrated a set of advantages. The results of its evaluation have been published in a prior work.

In the presence of a rich interaction environment, where interactions are related to programming teaching in higher education. We aim to expand the use of our IDE in order to improve this operation by harnessing the data that we can collect in it. The first step, is to define the learning analytics specific to this operation that will allow us to collect useful data. We believe that after processing this data we will be able to exploit it to improve the operation of teaching and learning.

The purpose of this work is to define a road map of the data that can be collected from web IDEs used in teaching as well as the possibilities of exploiting them.

Keywords: learning analytics, computer science teaching, E-learning 2.0.

1 INTRODUCTION

Currently, to support technology-based teaching and learning, it is not enough to make it available or to force students to use it, because it does not necessarily guarantee success. J. A. Larusson & al. [1] wonder “how do we effectively explore e-learning communities in order to get an accurate description of the complex interactions taking place inside virtual environments? What are the available analysis methods? How can the information collected enrich students’ learning experiences? How can we have a positive impact on teaching pedagogy? How to design even successful implementations of educational technology that report rich data to affect subsequent implementations? These questions are those needed to enhance educational program not only for “teaching with technology”, but simply for teaching in the first place”.

These processes, whatever their form, are quite complex. Teaching and learning can take place in a classroom, but can also take place out of class. Precisely, producing huge volumes of data where the vision of what data to collect, how to collect, and how exploring is not necessarily clear. In recent years, learning analysis (LA) has become an area that seeks to provide answers to questions such as those highlighted above.

The main objective of this work is to be able to exploit the information exchanged or generated by learning and teaching environments. Precisely, we have an environment that allows teaching and
learning of computer science practical work. We aim to exploit it to process Learning Analytics to improve the operations supported inside.

We begin this paper by presenting our virtual teaching and learning environment and its features. Then we introduce the concept of Learning Analytics, the environments and the examples of exploitation of this concept. Finally we present an achievement proposal to process data exploitation in our environment.

2 RELATED WORK

As part of the research we are conducting, which revolve improvement of computer-assisted teaching (Blended learning in particular), we have built a social Integrated Development Environment (IDE) called IDE 2.0 [2]. This IDE was designed based on Web 2.0 principles and it specifically propose the realization of computer science practical work or software projects in a code-centric environment. The following diagram illustrates the principles on which we have based our Development Environment:

Blended learning; consist of using online information and communication technologies as complement of traditional approaches within the institution. this phenomenon is increasing more and more and the use of technology in teaching becomes unavoidable [3]. The main objective of developing our environment is to have a system that make the realization of practical work easier. Which leads to a blended learning approach.

Collaboration: a serious requirement during programming work whether in the institutional context or not. We provided a particular care to the collaboration during the realization of the practical work of programming. Therefore, collaboration between the teacher and the student as well as the students themselves has become more fluent.

Real time editing: this is an emerging aspect that helps even more in collaboration. Collaborative work needs eventually editing the same resources in real time. Therefore, in source code editing context, the teacher and the student or the students themselves need to work remotely at the same time on the same code. Real-time editing has demonstrated prodigious efficiency in IDE 2.0.

Social Aspect: Through the advent of the social web, applications tend to implement more social interactions making platforms more dynamic and interactive. In IDE 2.0 we have used mush up techniques to integrate social network interactions into IDE 2.0 [4]. As a result, communication, comments, sharing, expressions are present on IDE 2.0 rendering it as social IDE.
3 LEARNING ANALYTICS

3.1 Definition and Objectives

Learning analytics LA is generally defined as collection, analysis and presentation of accumulated learner data to evaluate the behavior of educational communities, it is also defined in a broad sense as the effort to improve teaching and learning.

Practically, it is a predictive modeling, interaction visualizations or taxonomies and frameworks. Its objective is to optimize the performance of students and teachers, to refine teaching strategies, to rationalize institutional costs, to highlight potentially struggling students and refine the scoring systems in real-time [1].

Learning analytics encompasses a range of technologies, methods, models, techniques, algorithms and pedagogical best practices to understand and improve the reality of a student's learning. The following figure describes a reference model for LA based on four dimensions and identifying various challenges and research opportunities in the area of LA in relation to each dimension [5].

This model can be used to build a road map to build Learning Analytics and to exploit data and interactions in learning environments. In the present work we have been inspired by this reference model to define the possibilities of realization and exploitation of LA in our environment. Thereafter, we present the environments in which LA studies are already in progress as well as examples of studies.

3.2 Learning analytics environments

There are categories of students in the educational environment that transform the majority of their educational activities into interactions with digital systems. Moodle is an example of these environments that is implemented in an interesting number of institutions. This online learning platform is considered as an LMS (Learning Management Systems) as well as others like SAKAI and E-front. LMS supports online collaboration, tutoring systems and even administrative functions [6].

MOOCs "massive open online course" is an open type of distance learning able to accommodate a large number of participants [6]. A. McAuley & al. consider it as an online phenomenon gathering momentum over the last decade, a MOOC integrates the connectivity of social networking, the facilitation of an acknowledged expert in a field of study, and a collection of freely accessible online resources. Perhaps most importantly, however, a MOOC builds on the active engagement of several hundred to several thousand "students" who self-organize their participation according to learning goals, prior knowledge and skills, and common interests [8].
Some other tools are used as learning environments even if they are not designed specifically for that for example Wikis, social networks and video casts. Their use in learning has been successful due to the ease of content creation and publication and student familiarization with these environments.

In this context the student can learn one or more skills through these environments, he can also have access to a wide range of free educational resources that may or may not collect or transmit data. As a result, the use of many online activities would likely have positive learning outcomes. These learning activities can carry a very large and very rich amount of information that can be exploited. In the following section, we present examples of exploitation that explain the value of our contribution.

3.3 Experiences

In this part, we present some experiments that have been done for the same purpose of our work. Therefore, we are trying here to present the state of the art of the experiments already carried out which have shown preliminary results.

These experiences have common principles with our project which is the collection and the systematic analysis of learner's data making possible to contribute to their evolution. This helps as the process progresses over time, allowing institutions to use existing data to identify and interpret learners' difficulties in order to increase their success.

3.3.1 Experience 1

Purdue University has developed a system called "Signals" in 2007 [1] which examines the extent to which data derived from the LMS can be used to predict student performance. The challenge was to identify the student at risk of doing poorly in a classroom using only readily available data, namely, current course notes, information on the educational background of demographic descriptors, and data indicating to what extent a student interacted with the LMS. The goal then was to help students understand both their current classroom grades and what they can do to have a higher grade while there is still time to act [9]. More specifically, the goal was to inform students about the specific actions they could take to have a positive influence on their ranking in a course. This information was to be written by the instructor and transmitted via short e-mail messages and publications on the course's LMS site [9]. The novelty of "Signals", compared to other early warning systems, appears in the fact that it took into account students' efforts in measuring the degree of interaction of a student with the system.

3.3.2 Experience 2

Advisers at the STEM academy "Science, Technology, Engineering, and Mathematics" relied on self-reported student grades, which students brought to monthly meetings. But, the schedule of monthly meetings did not provide enough frequent interactions between students and counselors. For example, when a student failed an exam or task, it was often too late to correct the student's academic background. To solve this problem, they used "EWS" Early warning system developed to identify students in need of support and increase the frequency with which a counselor contacts them.

By collecting two generic data sources from the learning management system: quality data and connection data. The note data was collected by downloading the gradebook and assessment data from each course site. Login events were collected by querying the LMS data repository and counting the number of times a student accessed the site of a specific course. Notes and connection events have been aggregated and translated into a variety of visualizations, including numbers displaying students' scores relative to their peers over time, as well as lists of performance on entries. This information has been updated and sent to advisors weekly. They developed a three-level classification system: Engage (red), Explorer (yellow) and Encourage (green), providing advisors with a simple description of the complex relationships between academic performance data, including longitudinal data, intra-courses comparisons and connection events.

4 DATA EXPLOITATION IN TEACHING AND LEARNING CODING ENVIRONMENTS

The main objective of this study is to build a roadmap according to which we will define the various aspects that allow us to exploit data and interactions in the development environments used for computer science teaching and learning. In other words, we aim to define learning analytics specific to
this case to be able to improve the operation of teaching and learning in a technical and pedagogical level.

The first step with which we started is the implementation of a development environment-specific model based on the general reference model developed by chatti & al (figure. 1) [5]. The model specifies the definition of 4 essential elements (what? Who? Why? And how?).

4.1 What? (Data & Environment)

We plan to do learning analytics on our Development Environment which is a social, collaborative, learner-centered environment [2] (a kind of Virtual learning Environment (VLE) [10] specific to software development). We present in this part the dataset that can be used:

- **Technical data:** constitute the data relating to the technical skills targeted by the work carried out on the environment. The environment must allow all possible information in this category. For example, the system must present a student's data in relation to a skill in a specific course.
- **Pedagogical or didactical data:** the system must be able to provide relevant information concerning the didactics used to carry out practical work in our environment. For example: problem-based approach, project-based approach, demonstration-based approach …
- **Communication data:** our environment allows communication between different actors. This communication focuses on the operation of teaching and learning. We aim to exploit the communication data also between the teacher and the student or between the teachers or the students themselves. A very interesting aspect in communication is that it presents the emotions. This latter can be exploited in several ways.
- **Collaborative Data:** Collaboration is a crucial operation that helps in the operation of learning. Our environment has been created to facilitate collaboration and facilitate learning as a result. We aim to expand our system to collect data related to collaboration.

4.2 Who?

In this part we define the stakeholders of the environment. As defined in the general model [5] we take into account the students and teachers who present the main actors. Decision-makers and researchers are other essential actors for this environment. Decision-makers must be able to visualize information useful to make decisions and researchers (mainly in pedagogy) must be able to conduct research studies in the environment.

4.3 Why?

Here we must define the objectives for which we intend to exploit the power of the data. According to the basic model, we can exploit LA data for several reasons. Here we identify the same goals but applied in the case of development environments:

- **Monitoring/analysis:** the goal is to have mechanisms that can provide control of operations. An example of a fundamental need in relation to our system is the control of the educational and technical progress of teaching. Monitoring can be specific or global. The results at the global scale can be presented in the form of graphs and dashboards allowing decision-makers or researchers to extract useful information.
- **Prediction/intervention:** The exploitation of our data in Development Environments can benefit from predictive approaches to predict certain incidents related to teaching and learning operations. The goal is to apply a proactive or preventive approach to react before undesirable situations. A basic example of this goal is to predict students' cognitive progression over applied pedagogy. Predictions can lead to changing the pedagogical approach in order to have a better progression.
- **Tutoring / Mentoring:** the system must be able to guide the teacher or learner to follow a strategy for teaching to be effective. For example, the system can automatically adjust the volume of practice sessions according to the student's abilities and progress. The system can propose to a student to learn a prerequisite necessary to continue an ongoing learning.
• Assisted assessment: the objective is to rely on LA to build a LA-assisted assessment. An example may be the automatic proposal of the system to the student to pass an evaluation in case their old evaluations did not allow him to reach certain objectives.

4.4 How?
This section presents the techniques or technological tools that can be used. The model proposes as an example the following tools:

• Statistics & Information visualization: we can allow different actors in the environment to visualize collected information based on the statistics. Visualizing this information can help improve learning and teaching operations.

• Data Mining: These domain can help with its methods to present information in a more relevant way. This can help a better responsiveness and decision making.

• Social Network Analysis: Social network analysis emerged as an important research topic in sociology decades ago. The analysis was usually carried out as a “field study” on small communities, gathering data through questionnaires, interviews, and other labor-intensive methods [11]. In the presence of recent virtual social networks, this science has experienced a new phase. Our environment is full of social interactions. We can conduct social studies within our environment.

• Machine learning & deep learning: deep learning which is a Machine Learning discipline is the way in which the machine becomes able to learn automatically. It allows computer models composed of multiple processing layers to learn data representations with multiple levels of abstraction [12]. These methods have intensely improved the state of the art in speech recognition, visual object recognition, object detection, and many other areas. We rely on its techniques to enable the machine to interpret the content in our environment in an intelligent way.

• Big data: according to Chen, M & al. [13] big data is an abstract concept. Apart from masses of data, it also has some other features, which determine the difference between itself and “massive data” or “very big data”. In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. It is some time defined by its 4V feature (Value, Variety, Velocity and Volume) [14]. We also plan to explore this new area in order to use its techniques in our environment.

5 CONCLUSION
In order to improve institutional technology-based teaching and learning, this paper has addressed all aspects of learning analytics starting from its definition as a tool for collection and presentation of student’s learning data. The purpose is to evaluate their behavior in order to improve the process of teaching and learning.

We have also presented an overview of the most used environments in teaching and learning. LMS, MOOCs and web 2.0 tools are the most used. We then presented some examples of how learning analytics can be used to improve the process. This shows on one hand the importance of the exploitation of the learning data and in other hand this field promises a lot of results since it is still in the beginning.

By regarding the state of the art in Learning Analytics, we have found a basic model for LA that we have used as a guide to more clearly identifying the objectives, opportunities and techniques to use to exploit data and interactions in our environment.

We aim to implement the outcomes obtained in IDE 2.0 which was designed based on the Web 2.0 principles and which allows the realization of the practical work of computer science in a social and collaborative environment. We believe that this promises significant results as IDE 2.0 allows for full interaction in learning and teaching operations.
REFERENCES


