Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations

Riccardo MAZZA, Christian MILANI

Faculty of Communication Sciences, University of Lugano, Switzerland
{mazzar, milanic}@lu.unisi.ch

Abstract. This paper presents a novel approach of exploring usage analysis in learning systems by means of graphical representations. Learning systems collect large amounts of student data that can be used by instructors to become aware of what is happening in distance learning classes. Instead of being processed with techniques from user modelling, data is displayed "as it is". Techniques from Information Visualization show how useful insights can be gained from graphical representations. A system called GISMO illustrates the proposed approach. By presenting graphical representations, GISMO allows the user to visualize data from courses collected in real settings. We will show how using graphical representations of student tracking data enables instructors to identify tendencies in their classes, or to quickly discover individuals who need special attention.

1 Introduction

Today, web-based learning environments are widely used among universities around the world. They take advantage of the client-server communication on the Internet to provide instructors with a learning environment where they can distribute information to students, produce content material, prepare assignments and tests, engage in discussions, and manage distance classes [10].

In most cases, these tools are no more than sophisticated web server applications that help students and instructors in their learning activities. Even though such systems have been used in the last decade, they lack the latest functions that we can find in research prototypes. In particular, current commercial applications lack the possibility of understanding the users' needs and of adapting the content and presentation to a specific learner (adaptivity).

In a face-to-face classroom lecture, the teacher is able to perceive users' feedback, and he/she adapts the teaching according to the learners' comprehension. In a distance learning setting, because of the nature of computer-mediated communication, the tutoring lacks some specific modalities of interaction such as gestures, facial expression, direct dialogue, etc. This leads to one of the most common problems of distance learning from the tutor's perspective: the monitoring and checking of students' activities in courses delivered with distance education tools [4].

The interaction mediated by the digital media makes it difficult for the instructor to verify elements essential in didactic. There is a lack of understanding which part of the course an individual student or a group of students is working on, or the level of mastery achieved by each student for specific concepts of the course, etc.

Activities such as answering questions, monitoring and promoting discussions, monitoring the learners' progress, and testing the acquired knowledge and skills on a regular basis are essential for a good on-line tutoring practice [4][2][11]. While web-based

learning environments are supposed to help tutors to accomplish these tasks, they often provide complex, confused, and useless information [6].

This research attempts to bring some advanced features to the web-based learning environments with respect to adaptivity. Particularly, we are interested in giving instructors additional functionalities to help them in their teaching activities, and adapt teaching according to individual and class activities and progress. A prerequisite for this is the instructor's awareness of what is happening to his students: Are they reading materials? Are they regularly accessing the course? Do they engage in discussions? Are there quizzes or assignments particularly problematic to the students? Are their submissions performed in due time?

Most learning environments accumulate large data logs of the students' activities, and usually provide some monitoring features to enable instructors to view some aspects of the data: e.g. the history of pages visited, the number of messages read and posted in discussions, etc. Student tracking data provided by the learning environment is a valuable source of data that can be used by the instructor not only to check students' activities, but also to improve the quality of the materials. For instance, an instructor may check which part of the course materials are most or least accessed by the student. The instructor may then perform further investigations to understand whether the students found these parts difficult to understand or not. However, student tracking data is complex and is usually organized in some form of a tabular format, which in most cases is difficult to follow and inappropriate for the instructors' needs [6]. As a result, web log data is used only by skilled and technically powerful distance learning instructors.

2 GISMO – a Graphical Interactive Student Monitoring System

In order to explore an alternative method to represent student usage data, we implemented a tool that we called GISMO. GISMO uses the students' tracking data as source data, and generates graphical representations that can be explored and manipulated by course instructors to examine social, cognitive, and behavioural aspects of distance students. It implements some of the visualizations found useful by teachers, based on our experience with the CourseVis research [7][9], within a new context, namely the Edukalibre project funded by the European Union.

GISMO uses techniques from Information Visualization [1] to build graphical representations that an instructor can manipulate, which may help him/her to gain an understanding of his/her students and become aware of what is happening in distance classes.

From a technical point of view, GISMO is an application that runs in conjunction with a web-based learning platform, and it is delivered through the Web using a Java Applet. We considered the Moodle learning platform in this work for his Free and Open Source nature. However, it can be adapted to support other learning platforms. In fact, a software API is committed to retrieve some data that is usually present in a wide range of platforms such as moodle, claroline, fle3, mimerdesk, etc. as the data points are: discussions, accesses to the course, marks students receive in quizzes, and so on. This API can be adopted to support other platforms.

Figure 1 represents the welcome page of GISMO. As you can see, there are 3 different areas in the user interface:
1. *Graph Panel*: graphs are drawn on this panel.
2. *List Panel*: contains a list of students, groups, resources, quizzes, and assignments of the monitored course. For each list the instructor can select/deselect data to visualize.
3. *Time Panel*: using this panel the instructor can reduce the selection on time and

restrict the graph to a specific range of dates.

A copy of GISMO has been installed in the local Moodle platform at the University of Lugano. This installation manages about 250 courses and more than 2.400 users. It provides a valuable source of data that can be used to analyze GISMO's graphical representations in real settings.
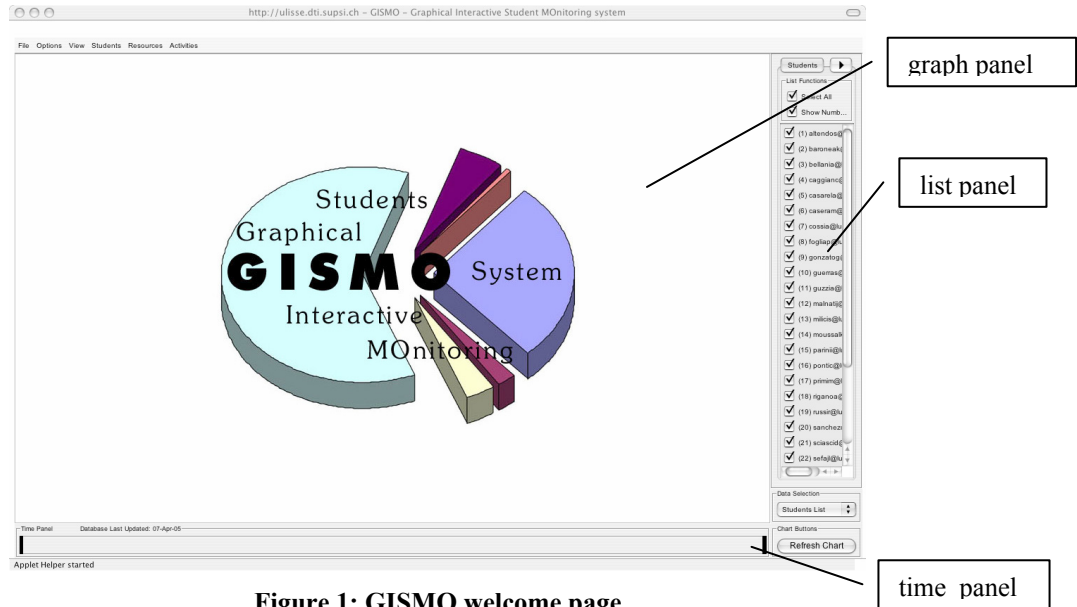

Figure 1: GISMO welcome page

In the next section we will illustrate some graphical representations of GISMO's abilities on data collected from real courses, and we will describe some insights that can be derived from representations. Representations were produced for the information regarded as interesting for instructors, that we had detected with a survey submitted to instructors involved in distance learning in a previous research [8]. That information is student attendance, access to resources, overview of discussions, and results on assignments and quizzes.
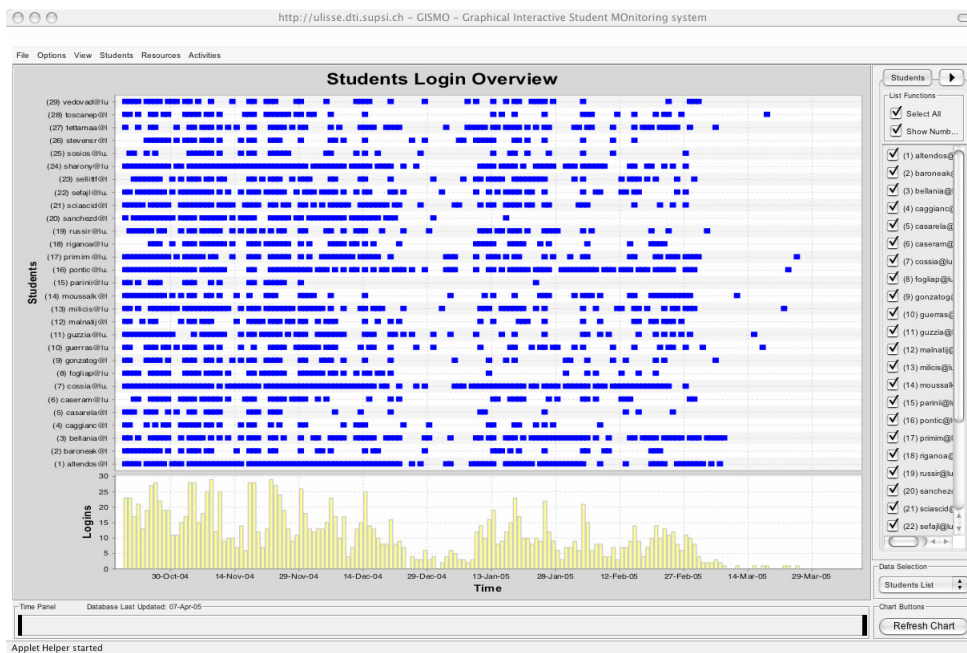
3 Students' attendance and reading of materials

Figure 2: A graph reporting the students' accesses to the course.

Figure 2 reports a graph on the students' accesses to the course. A simple matrix formed by students' names (on Y-axis) and dates of the course (on X-axis) is used to represent the course accesses. A corresponding mark represents at least one access to the course made by the student on the selected date. The histogram on the bottom shows the global number of hits to the course made by students on each date. With these graphs, the instructor has an overview, at a glance, of the global accesses made by students to the course with a clear identification of patterns and trends, as well as information about the attendance of a specific student of the course. For instance, we can clearly see how the attendance of the students during the first period of the course was particularly regular and uniform, with an inactive period at about the half way mark of the course (which corresponds to the Christmas period). It is interesting to notice how the accesses to the course become scattered during the second half of the course. The same image may allow the instructor to focus on about five students who were particularly persistent in accessing the course throughout (even in the second half of the course).
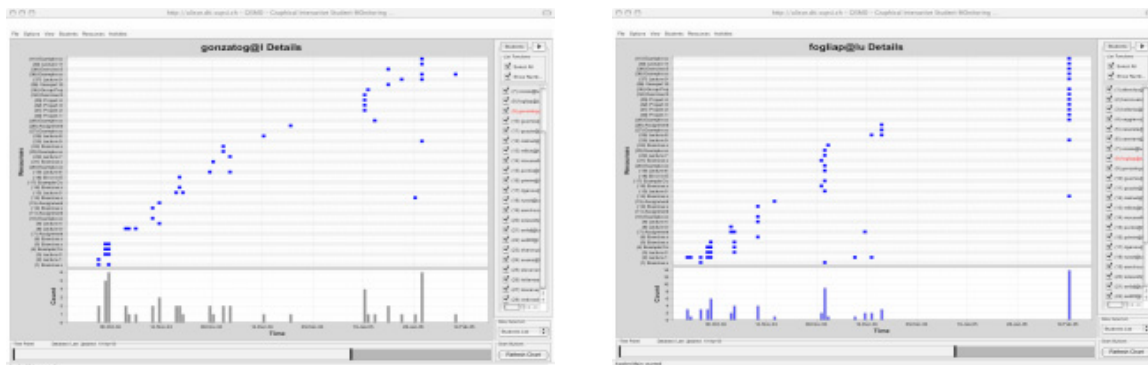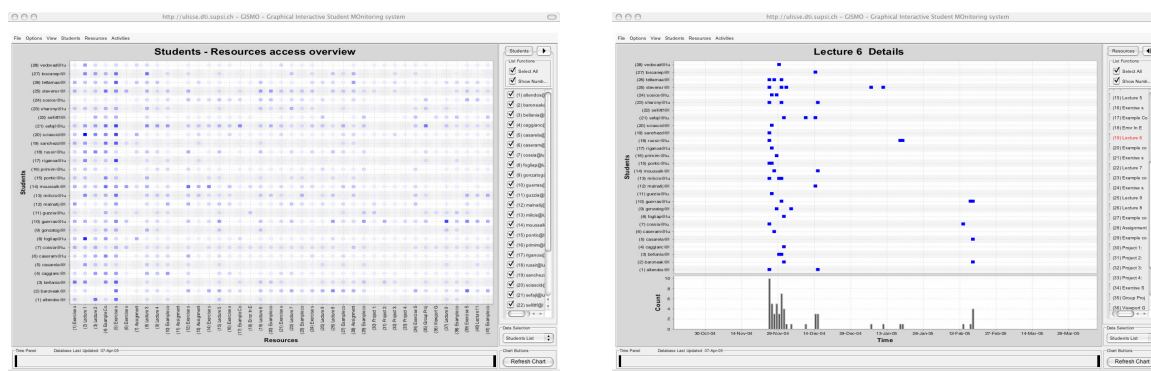

Figure 3: Two graphs reporting on the accesses performed by two different students to the course's resources.

Figure 3 reports an overview of the accesses of a student on the course's resources. By resources we mean any type of content that can be inserted into the course, such as a text

page, an assignment, an external link, a Power Point file, etc. Dates are represented on the X-axis; resources are represented on the Y-axis. Resources order on Y-axis reflects the resource sequence order inside the course. The histogram on the bottom represents the total number of accesses made by the student to all course's resources. Figure 3 shows two particular students of the course with different behaviors in accessing course resources. The graph on the left shows that this student regularly accessed materials. The graph on the right depicts a student having a different behavior: he/she concentrated his study during three periods: at the beginning, on a range of dates on the middle, and at the end (after a long time of inactivity he/she accessed several pages at once on the same day).

4 Resources

Figure 4: Two graphs reporting the overview of students' accesses to resources of the course (left), and the students' accesses to a particular resource of the course.

Instructors could also be interested in having the details on what resources were accessed by all the students and when. Figure 4 is intended to provide this information. The figure on the left reports student names on the Y-axis, and resource names on the X-axis. A mark is depicted if the student accessed this resource, and the color of the mark ranges from light-blue to dark-blue according to the number of times he/she accessed this resource. Some interesting insights can be seen. For instance, there are some resources intensely accessed by students of the course on the leftmost part of the graph (if the user puts the cursor of the mouse up one of the marks, a little tool tip appears showing the number of accesses made by the student on a particular resource). Moreover, there are some resources that had few accesses; these are easily identified by columns of the graph having few or no marks. Here the color propriety is used to provide a visual distinction between resources that had few accesses between those that had several accesses. This indication could be useful to the instructor to analyze the level of usage of the course material.

The Figure 4 on the right shows on which days the students visited a specific resource (with the graph on top) and how many global accesses they made to this resource for each day of the course (with the bar chart on the bottom). Again, student names are on the Y-axis, and dates are on the X-axis. This image can provide some insights to the instructor interested in knowing when a particular resource has been accessed during the distribution of time of the course.
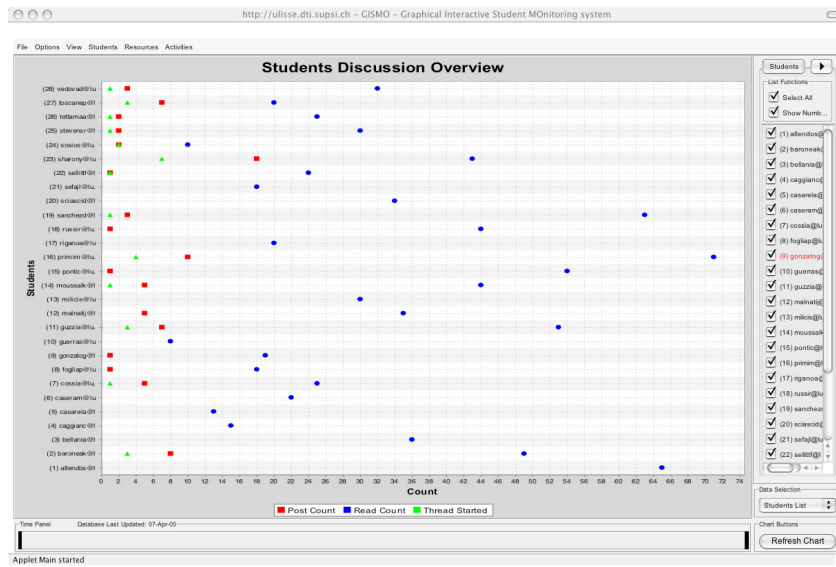
5 Discussions

Figure 5: Graphical representation of discussions performed in a course.

Discussions are a form of social activity that several instructors consider crucial in their courses [8]. Participation in discussion boards has to be considered for a comprehensive student analysis. The discussion board is a tool that allows students to read and post messages. Each message has a sender, a date, and a topic. A set of posts on the discussions, composed of an initial post about a topic and all responses to it is called a *thread*. Discussion boards' data is mapped onto a 2-D scatterplot and the generated image is illustrated in Figure 5. In this chart, instructors have an overview of all the discussions in which students participated. For each student of the course the chart indicates the number of messages posted (with a square), number of messages read (with a circle) and finally the number of threads started by the student in the discussions (with the triangle). It can be seen in the example in the figure that the most activity represented is reading, and less activity has been done in starting new threads or replying to other messages.
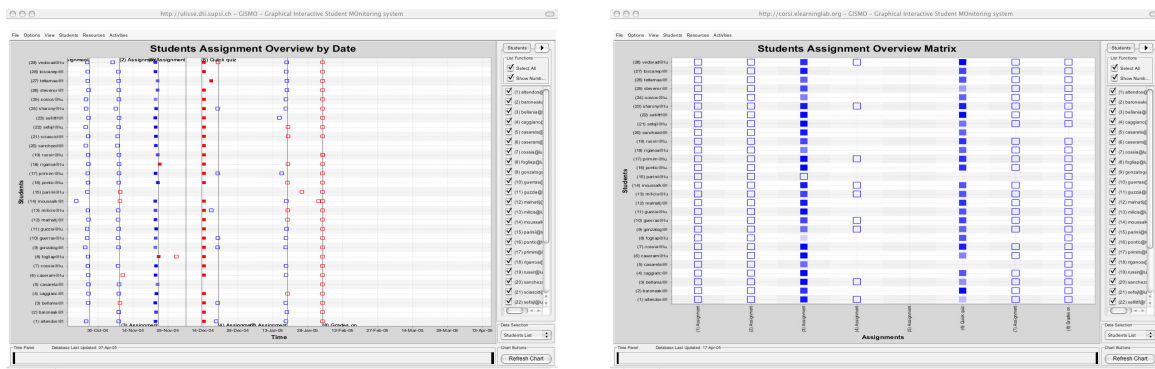
6 Assignments and quizzes

Figure 6: Two graphs reporting data from the evaluation tools. On the left instructors can see when the students submitted the assignment/quiz. On the right there is an indication of who submitted the assignment/quiz and an approximate indication on the grade. Different graphs are provided for quizzes and assignments.

Learning environments provide instructors with some tools to measure the level of comprehension achieved by students in the course's concepts. These tools are quizzes and assignments. For these two tools we produced the representations depicted in Figure 6. The tools collect data from these submissions: date and time of the submission, and, if this is available, the grade. This information is precious to the instructor who has to analyse the feedback on level of comprehension achieved by students, and then tune the teaching or the material accordingly. The left graph in Figure 6 is dedicated to visually indicate the date of submissions. Vertical lines correspond to deadlines of each quiz or assignment provided to students (represented here on the Y-axis). In this example it can be clearly seen that almost all submissions were made just in time when the deadline was almost approaching. Also, all students submitted their work late for the fifth assignment. Very few students submitted their solutions for the sixth assignment. Lines and marks have different colours for different quizzes or assignments to help the reader locate the marks corresponding to each.

Together with the submission time, the grade is another useful piece of information provided with the graph on the right. On the X-axis we mapped the assignments (or quizzes in the graphs dedicated to quizzes) and marks denote students' submissions. An empty square means a submission not graded, while a coloured square reports the grade: a lower grade is depicted with a light colour, a high grade is depicted with a dark colour. In the example, only the third and sixth assignment have been graded.

The graph allows micro and macro analysis of the students' performance. At the macro level, it can bee seen how most students submitted their solutions to each work, excepting the fourth and the fifth ones. At the micro level, the instructor may use the graph to detect problematic topics or students by comparing columns and rows. The instructor can also see how it is the performance of a particular student on a specific topic.

7 Conclusions and future work

We have presented a novel approach of using graphical representations of student tracking data collected by learning tools to help instructors become aware of what is happening in distance learning classes. A system, called GISMO, illustrated the proposed approach. GISMO has been implemented based on our previous experience with the CourseVis research, and proposes some graphical representations that can be useful to gain some insights on the students of the course.

Some forms of graphical representations have been explored in other works. Particularly, some forms of visualizing cognitive aspects of students have been explored in open student modelling projects, e.g. ViSMod [12] uses concept maps to render a Bayesian student model; UM [5] uses different types of geometric forms to represent known/unknown concepts; KERMIT [3] uses histograms to represent levels of a student's knowledge. The pictorial representations provided by such systems externalise a student model built by the system based on some Artificial Intelligence inference. Extracting student and group models can be fairly challenging, especially when dealing with large numbers of students. By contrast, graphical representations provided by GISMO merely represent data collected by CMS in a visual format with minimum data processing. In this case models are inferred in the instructor's mind, instead of being inferred by algorithms.

The GISMO is part of the research project "*EDUKALIBRE, Libre software methods for E-Education*" funded by the European Union in the years 2003 – 2005. This project aims at the translation of the uses and procedures of libre software (free/open source software) to the creation of content suitable to be used as material for education. Within this project, a pool of experts is currently evaluating GISMO with respect to the usability and pedagogical evaluations.

Acknowledgements

References

[1]   Card, K. S., Mackinlay, J. D., Shneiderman, B., 1999. *Readings in Information Visualization, using vision to think.* Morgan Kaufmann, Cal. USA.

[2]   Cotton, K. (1988). *Monitoring student learning in the classroom.* School Improvement Research Series (SIRS). Northwest Regional Educational Laboratory, U.S. Department of Education.

[3]   Danita Hartley, Antonija Mitrovic. *Supporting Learning by Opening the Student Model.* Proceedings of ITS 2002, Biarritz.

[4]   Helic, D. and Maurer, H. and Scherbakov, N. (2000). *Web based training: What do we expect from the system.* In Proceedings of ICCE 2000, Taiwan, pages 1689–1694.

[5]   J. Kay. *The UM toolkit for cooperative user modelling.* User modeling and user adapted interaction journal. Vol. 4, 1995.

[6]   Mazza, R. *Using information visualisation to facilitate instructors in web-based distance learning.* Ph.D. thesis, University of Lugano, 2004.

[7]   Mazza R., Dimitrova V. *Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education.* 13th International World Wide Web Conference (WWW 2004) - Educational Track. 2004 May 17-22, New York.

[8]   Mazza R., Dimitrova V. *Informing The Design of a Course Data Visualisator: an Empirical Study.* 5th International Conference on New Educational Environments. Lucerne 26-28 May 2003. (ICNEE 2003). pp. 215-220.

[9]   Mazza R., Dimitrova V. *CourseVis: Externalising Student Information to Facilitate Instructors in Distance Learning.* In U.Hoppe, F. Verdejo, J,. Kay (eds.) Proceedings of the International conference in Artificial Intelligence in Education. Sydney July 20-24, 2003. (AIED 2003). IOS press. pp. 279-286

[10]   McCormack, C. and Jones, D. *Building a Web-Based Education System.* Wiley, New York. 1997.

[11]   Ragan, L. C. (1998). *Good teaching is good teaching: an emerging set of guiding principles and practices for the design and development of distance education.* DEOSNEWS, The American Center for the Study of Distance Education, Pennsylvania State University, 8(12).

[12]   Diego Zapata-Rivera, Jim E. Greer. *Exploring Various Guidance Mechanism to Support Interaction with Inspectable Learner Models.* Proceedings of ITS 2002, Biarritz.