Development of a Data Mining Education Framework for Visualization of Data in Distance Learning Environments

3 authors, including:

Alexandre M. A. Maciel
Universidade de Pernambuco
66 PUBLICATIONS   **148** CITATIONS

SEE PROFILE

Rodrigo Lins Rodrigues
Universidade Federal Rural de Pernambuco
130 PUBLICATIONS   **355** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Avatar Educação: Desenvolvimento de um Ambiente de Learning Analytics para Apoio ao Processo de Ensino-Aprendizagem View project

Project   Avatar Educação: Um Assistente Virtual Inteligente Integrado ao Ambiente Virtual de Aprendizagem Moodle View project

Development of a Data Mining Education Framework for Data Visualization in Distance Learning Environments

Angelo F. Dias Gonçalves and Alexandre M. A. Maciel

University of Pernambuco - UPE
Recife, Brazil
{afdg, amam}@ecomp.poli.br

Rodrigo L. Rodrigues

Rural Federal University of Pernambuco
Recife, Brazil
rodrigo.linsrodrigues@ufrpe.br

Abstract— **With the increasing interest in developing Learning Analytics tools that can be integrated into the well-known Moodle course management systems nowadays, many tools have already been developed. These tools usually requires the user to know data mining techniques, and also requires time to get mining results from the tools. To address this problem, in this article, we present a structure that uses pre-built data mining through Shiny to quickly obtain results with a focus on visualizing data with graphs, and thus allows the integration of other research, called FMDEV. Guided by the proposed framework, a tool was developed for display the data mining results in a few clicks for Moodle users who wish to have them for day-to-day use and allows users with more experience in data mining to integrate new research. Finally, we used FMDEV tool to generate some experimental results using a set of real-life sample data from undergraduate students.**

Moodle; Distance Learning; Framework; Educational Data Mining; tools.

I. INTRODUCTION

The last Brazilian census for learning distance in higher education was published in 2016. It was realized with 339 institutions coming from all administrative categories, federal, state and municipal, public, private and non-profit, institutions of the National Learning Service (e.g., Sesi, Sebrae, Senac, Senai etc.), non-governmental organizations (NGOs), third sector and public agencies [1].

According to EaD.BR 2015 Census [1], most of the professionals working at Distance Education (DE) are tutors and teachers, in which 29,380 tutors and 18,769 teachers. In EaD.BR 2014 Census [2], 17,692 tutors and 11,074 teachers were raised. It represents a growth of almost 60%. These professionals are responsible for teaching courses that currently have 5,048,912 students in Brasil [1].

A Learning Management System (LMS) gathers vast amounts of information about students' interactions. Most of this information refers to communication through existing tools such as discussion forums, chats, and digital media postings. These tools, when used, generate information on how and when students execute their assignments, tasks, course commitments, etc. The Moodle's default reporting tool offers information and filtering capabilities, but the data they provide is considered raw data and does not provide meaningful information about the teaching-learning process [3].

According to Luna, Castro, and Romero [4], extracting this type of useful data and transforming such information into actionable knowledge is a difficult task, and Educational Data Mining (EDM) offers both different and convergent perspectives, methodologies, techniques, and tools aiming to facilitate the process of discovering knowledge from this environments. EDM is defined as the development of methods for exploring the unique types of data that come from educational settings and using those methods to understand students and what and how they learn [5].

Nevertheless, currently, although existing tools offer different levels of analysis, basic graphs on user interaction with the platform, and some data mining techniques, manipulation of these tools still requires prior knowledge of data mining, and is time-consuming to use day to day. Based on this, it is necessary to expand this information extracted from data mining to teachers and tutors from all areas.

To overcome this drawback, we developed a particular framework, called visual educational data mining framework (in Portuguese: FMDEV), which enables the extraction of knowledge and exports it visually to a larger number of LMS users. Finally, to extend the framework objective, we developed a tool for the purpose of validating the framework, because tools like this are more precise and easier to be used by non-expert users in data mining than general or traditional data mining tools [7]. In this way, instructors can discover, in a simple way, hidden information about students' behavior, and how students learn in the courses in a single and user-friendly interface [8].

The objective of this study is to present a valid framework that can help the main users of LMS Moodle in daily decision making, independent of data mining knowledge, thus providing a tool and its experimental results. Therefore, this article is

_____

organized as follows: Theoretical Foundation in Section II, Framework presentation in Section III, Validation of the Framework in Section IV, and finally, Conclusion and Future Works in Section V.

II. THEORETICAL FOUNDATION

A. Data Mining and Educational Data Mining

Data Mining (DM) can be considered as a central stage within the more general knowledge discovery process [6]. The objective of data mining is to determine models that include the relations among data and allow a better understanding, prediction or generalization of these data.

According to Fayyad, Shapiro and Smyth [9], the main idea of data mining techniques is feature selection and pattern recognition of database system. Data mining can be declared as the knowledge discovery from an enormous amount of raw data using statistical methods, machine learning, and artificial algorithms.

According to Romero and Ventura [5], Educational Data Mining is the field that discovers new knowledge based on students', teachers', tutors' and manager' usage of the system to improve the quality of the education offered through online courses. In their review of state of the art about EDM, they list eleven educational tasks that make use of data mining techniques.

Among these tasks we highlight two that are related to the objective of this work:

- 'Analysis and Data Visualization,' which aims to highlight useful information and support decision making.

- 'Providing Feedback for Supporting Instructors', which supports course authors/teachers/administrators and enables them decision making (about how to improve students' learning, organize instructional resources more efficiently, etc.) and allow them to take appropriate proactive and/or remedial action. It is important to point out that this task is different than data analyzing and visualizing functions, which only provide necessary information directly from data (reports, statistics, etc.).

B. Shiny and Sharing Apps

Shiny [10] allows the non-expert computer scientist to create publication-ready figures and tables through an intuitive interface to the underlying computer code.

The Shiny can build a web-based application which is published on the Internet. It is easy to develop and to integrate into a web content using HTML and CSS [11]. The Shiny is an R package which is responsive to display the results of the data mining analysis into web-based applications. This web-based application was developed using the R programming language which is open source with several attractive packages [12].

The Shiny has two structures, contained in the files server.R and ui.R. The server file is a set of instructions that build the R components while the user-interface file is a set of instructions to display the application. It is already possible to find in the literature several web-based applications using Shiny and with several techniques of data mining and algorithms [13] [14].

Sharing is the central part of Shiny applications. To publish your Shiny app, only organize the files and post to a server in the cloud. RStudio insists on this sharing by providing a free server, called shinyapps.io [15]. Shinyapps.io is a server maintained by RStudio, where it is possible to make online applications available for easy access and use. Through this server, it can access any shared application from anywhere.

III. FRAMEWORK

FMDEV tool has been developed in HTML5, CSS, and JavaScript language and integrated into Moodle as an HTML block. The main reason for integrating the proposed tool as an HTML block is because any of these HTML blocks provide the ability to create external links, which is desired by our tool.

Blocks are items which may be added to the left or right column, smaller sizes, or central, larger size, on any Moodle page. They contain information or new applications geared towards students or instructors. In our case, the added HTML block allows only the tutors/teachers to visualize the results exported by the framework and to have access to the administration of the framework through a tool button. Anyone responsible for administering LMS Moodle, before adding the block, must insert the framework package to the same server. In this way, it is easy to install FMDEV into existing Moodle deployments.

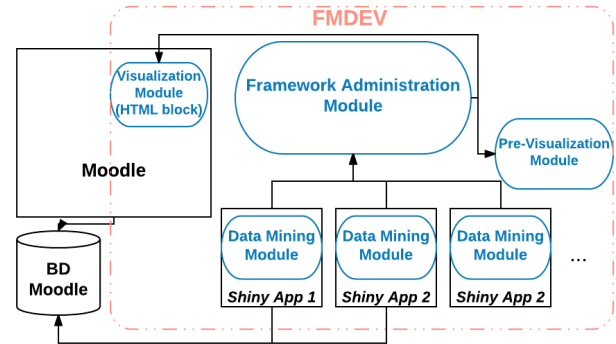From a high-level viewpoint, FMDEV comprises four main modules (see Fig. 1).

Figure 1. *FMDEV Architecture.*

FMDEV allows several data mining modules to be incorporated, where each module consists of a Shiny application, where they will be items in the administration list of the framework. Therefore, the proposed tool consists of three consecutive steps:

- The Pre-Visualization module enables the user to visualize the information that will also be shown in the visualization module and allows the user to return to the framework administration module, for a new selection, through the back button. It also allows the user to close the screen so that the user can return to Moodle.

- The visualization module, in the HTML block of Moodle, enables the display of the result of the first

module of data mining embedded in FMDEV. In this module, there is a button, in tool format, that when clicked the user is redirected to framework administration module.

- The Framework Administration Module enables the user to select one of the available data mining module options from the drop-down menu. When the user choice an option, he is redirected, in this step, to the pre-visualization module.

IV. VALIDATION

To validate FMDEV it was realized a benchmark with some found tools, based on the web, as well as an experiment with real data.

A. Benchmark

Currently, almost all new EDM tools are web based. We compare them with some examples next. Table 1 illustrates a comparison of the FMDEV with some web-based EDM tools.

This benchmark evaluated relevant points for our research, such as: Previous knowledge in data mining to visualize the results (Data M.); Allows inclusion of data mining projects beyond those already included in the package (Personal.); is it free? (Free); is it integrated to the LMS? (Integr.); and is it open source? (Open S.).

TABLE I.   BENCHMARK OF EDM WEB-BASED TOOLS.

Tool	Data M.	Personal.	Free	Integr.	Open S.
FMDEV	No	Yes	Yes	Yes	Yes
GISMO [16]	Yes	No	Yes	Yes	Yes
SNAPP [17]	Yes	No	Yes	Yes	No
AAT [18]	Yes	No	No	No	No
MOClog [19]	Yes	No	Yes	Yes	No
E-learningWebMiner [20]	Yes	No	No	No	No
CVLA [21]	Yes	No	No	Yes	No
IntelliBoard.net [22]	Yes	No	No	Yes	No
SmartKlass [23]	Yes	No	Yes	Yes	No
MEAP [24]	Yes	No	Yes	Yes	Yes
Analytics graphs [25]	Yes	No	Yes	Yes	No
VeLA [26]	Yes	No	No	Yes	No

B. Experiment

For the purpose of validating the operation of the proposed FMDEV, a real use experiment was realized. The LMS used was the Moodle of the Nucleus of Distance Learning (NEAD) of the University of Pernambuco (UPE), so the data used was gathered from on People Management discipline, part of 3rd-year reading Pedagogy course, with almost 200 enrolled students.

In the tests done two mining modules, the "shiny" apps, were developed, in which one of them was created based on the work of Machado [27], which used clusters analysis. The teacher / tutor will only need to select one of the modules as previously mentioned, but the administrator must be able to generate / integrate new data mining modules.

Fig. 2 shows FMDEV block already included into Moodle (on the right side), where it corresponds to the extracted view of the selected mining module, and the administration button (Tool button), which is in the lower right corner.
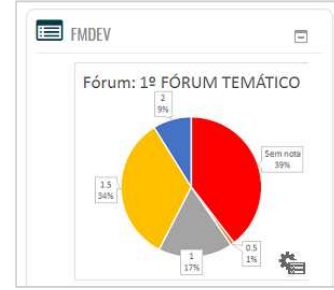

Figure 2.   FMDEV block in a real course.

The developed Moodle block, as previously stated, has been developed in HTML5, CSS, and JavaScript language. It's possible to access the Framework Administration Module, also implemented as an HTML5 page. The administration page of the framework could be hosted on a server different from the one found in Moodle, just as the mining modules, shiny applications, were hosted on a separate server.

A simple front-end has been implemented, so the user does not need more than two clicks to select the desired option. In this implementation, the Shiny app are presented as options, allowing the choice between one of them in a drop-down menu in the administration page of the framework.

Soon after selecting the desired option, the user is redirected to the pre-visualization module, in which the same one that will be shown on this screen will be updated in the Moodle block. Fig. 3 shows the pre-visualization screen as described above.
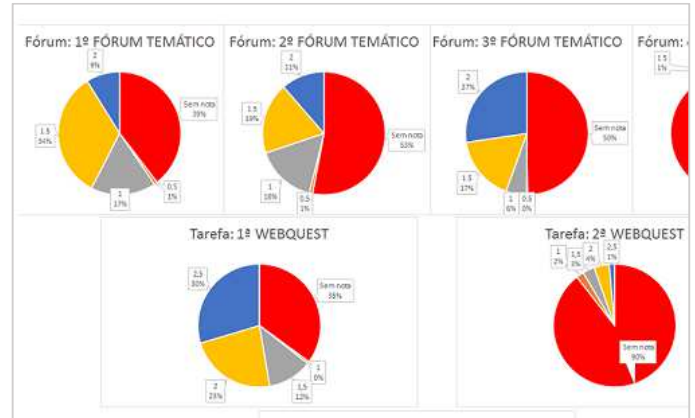

Figure 3.   FMDEV pre-visualization module.

Finally, what is shown on this page will also be used to update the Moodle HTML block, previously shown in this work in Fig. 2.

V. CONCLUSIONS AND FUTURE WORKS

Moodle management systems include a growing number of data mining tools that provide various types of analysis and graphing of user interaction. However, it 's hard to find any integrated tool that allows you to obtain results quickly and without the cost of time, so that it can be used in the day to day routine of tutors and teachers. In this way, the objective of this

article is to propose a new architecture, implemented that architecture in the form of FMDEV, which it´s used as a tool allows the discovery of knowledge in Moodle data.

Finally, the utility of the proposed tool is described with an experiment using as consolidated LMS, and a real-life dataset of Pedagogy students. This experiment mainly illustrates the ease of obtaining data visually and quickly as the user chooses. The results obtained provide useful information so that users can provide faster feedback on how students' progress in distance learning courses.

In the future work, we intend to implement FMDEV in some real course and monitor the use of the tutors/teachers, giving support and adding new modules of data mining. We are currently working on the development of an improved visual interface, adapted to the size of the screen, to present the results obtained without the need of a scroll bar. Also, we would also like to implement many modules for mining academic papers, and, not least, to register our tool with the National Institute of Industrial Property (INPI in Brazil is the federal body charged with executing and enforcing laws regarding the regulation of Any Industrial Property).

REFERENCES

[1] ABED, Associação Brasileira de Educação A Distância, "Censo EaD.br: relatório analítico da aprendizagem a distância no Brasil 2015", 1 ed. Curitiba: InterSaberes, 2016.

[2] ABED, Associação Brasileira de Educação A Distância, "Censo EaD.br: relatório analítico da aprendizagem a distância no Brasil 2013", 1 ed. Curitiba: Ibpex, 2014.

[3] M. A. Conde, A. Hernádez-García, F. J. García-Peñalvo, and M. L. Séin-Echaluce, "Exploring student interactions: Learning analytics tools for student tracking", International Conference on Learning and Collaboration Technologies, 2015.

[4] J. M. Luna, C. Castro, C. and Romero, "MDM tool: A data mining framework integrated into Moodle", Computer Applications in Engineering Education, Vol. 25, 2017.

[5] C. Romero, and S. Ventura, "Educational Data Mining: A Review of the State-of-the-Art", IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviews, Vol. 40, 2010.

[6] O. Maimon, L. and Rokach, "The Data Mining and Knowledge Discovery Handbook", 2nd ed., Springer, 2010.

[7] R. Mazza, M. Bettoni, M. Faré, L. and Mazzola, "MOCLog—Monitoring Online Courses with log data", 1st Moodle Research Conference Proceedings, Heraklion, Greece, 2012.

[8] S. Graf, C. Ives, N. Rahman, and A. Ferri, "AAT-A tool for accessing and analysing student's behaviour data in learning systems", In:Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Lak, Banff, AB, Canada, 2011.

[9] U. Fayyad, G. P.-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Magazine, 1996.

[10] Shiny by RStudio, "Easy web applications in R", Available from: https://www.rstudio.com/products/shiny/.

[11] C. Beeley, "Web application development with R using Shiny", Birmingham: Packt, 2013.

[12] W. Chang, J. Cheng, J. J. Allaire, Y. Xie, J. McPherson and M. Otto "Package 'shiny' ", 2015, Available from: http://shiny.rstudio.com.

[13] G. P. Siknun and I. S. Sitanggang , "Web-based classification application for forest fire data using the shiny framework and the C5.0 algorithm", 2016.

[14] R. Hermawati and I. S. Sitanggang, "Web-Based Clustering Application Using Shiny Framework and DBSCAN Algorithm for Hotspots Data in Peatland", Sumatra, 2016.

[15] Share your Shiny Applications Online, "Shinyapps.io by RStudio", Available from: http://www.shinyapps.io/.

[16] M. Luca, N. Mauro, M. Christian, and M. Riccardo, "L'aggiornamento del tool di monitoraggio delle attività degli studenti: GISMO 2.0", In: Atti del convegno italiano MoodleMoot, Bari, Italia, 2010.

[17] A. Bakharia, and S. Dawson, "SNAPP: A bird's-eye view of temporal participant interaction. International conference on learning analytics and knowledge", ACM New York, NY, USA, 2011.

[18] S. Graf, C. Ives, N. Rahman, and A. Ferri, "AAT-A tool for accessing and analysing student's behaviour data in learning systems", In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Lak, Banff, AB, Canada, 2011.

[19] R. Mazza, M. Bettoni, M. Faré, and L. Mazzola, "MOCLog—Monitoring Online Courses with log data", 1st Moodle Research Conference Proceedings, Heraklion, Greece, 2012.

[20] D. García-Saiz, and M. E. Zorrilla, "A service oriented architecture to provide data mining services for non-expert data miners", Decis Supp Syst 55, 2013.

[21] B. Dragulescu, M. Bucos, and R. Vasiu, "CVLA: Integrating Multiple Analytics Techniques in a Custom Moodle Report", International Conference ICIST, Druskininkai, Lithuania, 2015.

[22] IntelliBoard, "IntelliBoard.net", 2015, Available from: http://www.intelliboard.net/.

[23] Klassdata, "SmartKlass: The Learning Analytics Plugin", 2015, Available from: http://klassdata.com/smartklass-learning-analytics-plugin/.

[24] D.Y.-T. Liu, J. Froissard, D. Richards, and A. Atif, "An enhanced learning analytics plugin for moodle: Student engagement and personalised intervention", In: Australasian society for computers in learning and tertiary education (ascilite 2015), Perth, Australia, 2015.

[25] J. Singh, "New Block: Analytics Graphs", 2015, Available from: http://www.Moodleworld.com/2015/03/31/new-block-analytics-graphs.

[26] J. Cruz-Benito, and F. Garcia, "VeLA: A visual eLearning analytics tool", LASI, Bilbao, 2015.

[27] C. J. R. Machado, B. R. B. Lima, A. M. A. Maciel, and R. L. Rodrigues, "An investigation of students behavior in discussion forums using Educational Data Mining", International Conference on Software Engineering and Knowledge Engineering (SEKE), Redwood City, San Francisco Bay, California, 2016.