The data visualization technique in e-learning system

Eleonora Brtka*, Vladimir Brtka*, Visnja Ognjenovic* and Ivana Berkovic*
* University of Novi Sad/Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
norab@tfzr.uns.ac.rs, brtkav@gmail.com, visnjao@tfzr.uns.ac.rs, berki@sbb.rs

Abstract—Data mining techniques and their applications are widely recognized as powerful tools in various domains. Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational context. In the domain of education, there are varieties of data of different types collected during the educational process. The main question is: Is it possible to process the collected data with the data mining system and what are main advantages of data mining and e-learning interaction? In this paper, we present an insight into the possible interaction between course management system and data mining techniques. The main goal of this work is to investigate some data mining techniques in order to deliver most appropriate learning object to the learner. In this paper, visualization as a data mining technique is investigated. For this purpose, free data mining tool Weka was used.

I. INTRODUCTION

The changes that are constantly taking place in terms of rapid technical and technological developments affect society as a whole. Systems such as educational experience changes in terms of modernization. Students are required to improve the style of self-teaching and to improve their overall skills. In the improvement process the information capacities are very important, especially the resources available via the Internet. The using of the Internet and Course Management Systems (CMS) are not rare occurrence in Serbia but it still cannot be said that such systems are widely deployed. CMS can offer a great variety of channels and workspaces to facilitate information sharing and communication among participants in a course. They let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums, chats, file storage areas, news services, etc. These e-learning systems accumulate a vast amount of information, which is very valuable for analyzing students’ behavior, and could create a “gold mine” of educational data [1]. They can record any student activities involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers [2]. They normally provide a database that stores all the system’s information: personal information about the users (profile), academic results and users’ interaction data. However, due to the vast quantities of data these systems can generate daily, manual analysis is not possible. Instructors and course authors demand tools to assist them in this task, preferably on a continual basis. Although some platforms offer reporting tools, it becomes hard for a tutor to extract useful information when there are a great number of students [3]. They do not provide specific tools allowing educators to thoroughly track and assess all learners’ activities while evaluating the structure and contents of the course and its effectiveness for the learning process [4]. A very promising area for attaining this objective is the use of Data Mining (DM) [5]. Efforts that are being made in the integration of CMS and DM systems are evident. This integration often involves addition of the DM module to an existing CMS [6, 7]. The DM module and CMS are often connected in sequence [8]. Sequential or serial connection, in this case involves the collection of data by CMS, and then processing them with the DM module.

The structure of the paper is following: Section II contains basic notions about DM techniques and its possible application in e–learning system. Section III contains notions about visualization, as well as an illustrative example. Finally, Section IV contains conclusions and guidelines for future work.

II. DATA MINING TECHNIQUES

DM or Knowledge Discovery in Databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections [9]. DM is a multidisciplinary area in which several computing paradigms converge: decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming, statistical algorithms, etc. The most general and well-known DM techniques are [8, 10]:

- Clustering – determines the segmentation or data groups that are not predefined.
- Classification – classifies instances into predefined classes.
- Association rule mining – technique determines the data that can be classified into the same group. This technique is also known as dependency modeling.
- Description and visualization – visual (graphical) representation of the data.

Certainly, it is not always necessary to apply all listed DM techniques. The choice of technique depends on the type and volume of data to be analyzed. The guidelines in the selection of techniques and their detailed descriptions are available in [8].
A. Data mining tools

DM tools that are used to analyze data from different domains are:

- Rosetta allows the classification of data and generates decision rules in the If ... Then form. Researchers at the University of Warsaw and the University of Trondheim developed Rosetta software system. It is based on the application of the Rough sets theory [11].
- Weka system has been developed by researchers at the University of Waikato, New Zealand [12].

Rosetta can be used to induce and validate classification rules from a database. The data can arise in various ways (regular survey), and their form does not have to be adapted to data mining techniques. The system, which is most interesting for data visualization, in this case is the Weka system. This system is free, easy to use and powerful enough to visualize data collected by CMS or by some other means.

The Weka system allows:

- Classification (Discriminate analysis) that predicts class labels. This is supervised classification, which provides a collection of labeled (pre-classified) patterns [13].
- Clustering, data is divided into classes that are not previously defined [14].
- Generating association rules. This option discovers relationships among attributes in databases, producing If ... Then statements [15].
- Selection of the most relevant attributes (properties). Attribute selection involves searching through all possible combinations of attributes in order to find which subset of attributes works best for prediction.
- Visualization, view an interactive 2D plot of the data. Visualization is a branch of computer graphics and user interface, which is concerned with the presentation of interactive or animated digital images so that users can understand data [16]. This technique facilitates analysis of large amount of information by representing the data in some visual display.

Rosetta and Weka systems are particularly suitable for data analysis in the field of education, offering choice of DM techniques and are relatively simple to use.

B. Data mining and the CMS

There are many general and specific data mining tools [8]. Some commercial mining tools are DBMiner, SPSS Clementine and DB2 Intelligent Miner, while some public domain mining tools are Weka and Keel.

If there is a need to establish a serial connection between CMS and DM system in order to investigate data collected by CMS, certain steps must be undertaken. These steps are:

1. Selection and installation of CMS. This step requires the approval of management as well as definition of the hardware – software platform to support CMS.
2. Data collection by CMS. It is appropriate that the data are available in the form of MS Excel spreadsheets because that format is widely recognized and therefore supported by many DM systems. CMS frequently offer the option "Export" data to the desired format.
3. Data analysis by DM tools. In this step, one or more DM tools do analysis.

Recently, visualization technique is recognized as a useful tool for data analysis in the domain of education. This is particularly useful when log file produced by CMS is available. This kind of log file contains various information about CMS users.

At the Institute of Applied Pedagogy and Psychology of the Budapest University of Technology and Economics, with the active participation of the Centre for Teacher Training and Engineering Education of the Budapest, the investigation of electronic syllabuses is conducted [18, 19]. The main objective of the investigation was to explore the most important characteristics of student during the accomplishment of the Moodle courses. Web mining methods were used in this investigation. Parameters such as number of visits to a certain learning object and spent time were monitored and analyzed. Analysis was conducted by SPSS Clementine and Google Analytics and afterwards it was decided what amendments should be included in the syllabus.

Another example is available in [7]. Here, global number of accesses to the CMS made by students (in X-axis) to all the resources of the course (Y-axis) is graphically presented. Using the graph of this kind, the instructor has an overview of the global access made by students to the course. Clear identification of patterns and trends, as well as information about the attendance of a specific student in the course is possible. This information is very useful because instructor can more easily detect students with some learning problems or to define the changes that need to be entered into the course content.

Possible connection between CMS and DM is the “loop cycle”, see Fig. 1. DM analyzes the feedback from CMS; results of analysis are transformed to the guidelines for content changes. This is one way to achieve adaptability of the CMS.

![Figure 1. CMS – DM connection](image-url)
Visualization allows establishing different relationships and finding new ways of connecting information. It is possible to assess how to present information in a more meaningful way and get students more engaged.

III. VISUALIZATION

Visualization is a field of Computer science that examines techniques for pictorial representation of vast amount of abstract data, so that the data can be comprehended and interpreted by people. Visualization is a powerful method that comprises of three major tasks:

• Exploration (search for relationships, trends, and interesting phenomena).
• Confirmation (validation or refuting hypotheses).
• Presentation (conveying information to others).

In its scope, visualization is not limited to displaying only data. It is possible to graphically display articles, resources, connections, news, tools, services, web sites, even mind maps. Some of the visualization algorithms and techniques are volume visualization, information visualization, multiresolution methods, modeling techniques and interaction techniques and architectures.

By managing complex multi–dimensional data with appropriate visualization techniques people form mental models of the data and obtain a better understanding of specific features of the data [16]. Data visualization is a branch of computer graphics and user interface, which is concerned with the presentation of interactive or animated digital images so that users can understand data.

These techniques facilitate analysis of large amounts of information by representing the data in some visual display. Normally large quantities of raw instance data are represented or plotted as spreadsheet charts, scatter plots and 3D representations.

Data visualization can be used to graphically render complex, multidimensional student tracking data collected by web–based educational systems. The visualized information may be related to complementary assignments, admitted questions, exam scores, etc. [17].

Therefore, it can be expected that suitable pictorial representations of data from CMS will help the instructors to form mental models of individual students as well as mental models of groups of students.

By using these models, the instructors can provide instruction that is more effective.

A. An Illustrative Example

The goal of the data visualization is to help to identify appropriate learning objects. Data were generated by Moodle system and by surveying students.

A total of 256 instances (students) were processed. After using the CMS, students are asked to fill in an electronic questionnaire about usefulness of learning materials presented by CMS. The intersection between this information and the information on whether the student engaged CMS during the school hours or from home (own a computer) is important to distinct learning styles during the teaching and while learning at home.

DM technique that was used is visualization and a tool is Weka. Weka’s visualization section allows you to visualize 2D plots of the current relation (see Fig. 2). When you select the Visualize panel, it shows a scatter plot matrix for all the attributes, colour coded according to the currently selected class. It is also possible to change the attribute used to colour the plots, to select only a subset of attributes for inclusion in the scatter plot matrix, and to sub sample the data.

By clicking on a cell in the scatter plot matrix, a visualization of the selected scatter plot is displayed in a separate window. The data points are plotted in the main area of the window. At the top, there are two drop-down list buttons for selecting the axes to plot. The one on the left shows which attribute is used for the x-axis; the one on the right shows, which is used for the y-axis. Beneath the x-axis selector there is a drop-down list for choosing the colour scheme. This allows you to colour the points based on the attribute selected.

After pressing the Visualize All button, histograms for all the attributes in the data are shown in a separate window (see Fig. 3.).

Data visualization provides the "picture" of data, so it can visually observe the rules, which may exist. Figure 3. shows the influence of attributes: The computer helps ("yes" is represented in blue, "not" is represented in red) on the value of other attributes.

It may be noted that 210 students are using CMS from home (own a computer), while 46 students are not. It is possible to see visually how many students who are using CMS consider that computer is not useful (computer helps), or how many students who do not use CMS from home consider that computer is useful. In addition to this by monitoring student’s access to certain learning objects during the school hours or from home (number of access and duration), it is possible to make a distinction between learning styles during the teaching and while learning at home.
Analogous analysis can be done in the case of other attributes. The value of data visualization is that it allows the identification of dependence in a “single view”.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, a visualization section of the Weka data mining tool had been used for analysis of data gathered by CMS and by questionnaire. Various parameters were monitored, but in this work two parameters are crucial: student position on whether the CMS is helpful (yes, no) and whether a student uses a CMS during the school hours or from home. In addition, the number of visits to the certain learning object, as well as visit duration are used to differentiate between learning stiles during the school hours or at home.

Further conclusion had been drawn: in a given amount of time the number of visited learning objects was bigger when student used CMS during school hours, while duration of the visit was bigger when student used CMS from home.

Furthermore, it was possible to notice which learning material were used during school hours and which learning material were used while student was at home.

This allowed the administrator to change the priority of certain learning objects (depending on where they are accessed) for students, thus achieving a certain degree of adaptability of the system. The adaptability of the system leads to implementation of different and specific pedagogical strategies, which is in line with results presented in [20].

Future work will include more attributes (parameters) that are monitored. An important task will be to consider the options for greater adaptability of the system, as well as the automation of the adaptation process.

ACKNOWLEDGMENT

Supported by the Ministry of Science and Technological Development, Republic of Serbia, under the project number TR32044 “The development of software tools for business process analysis and improvement”.

REFERENCES