Integrating a Web-based ITS with DM tools for Providing Learning Path Optimization and Visual Analytics

Igor Jugo Božidar Kovačić Vanja Slavuj
Department of Informatics
University of Rijeka,
Radmile Matejčić 2, Rijeka, Croatia
+38551584711
ijugo@inf.uniri.hr bkovacic@inf.uniri.hr vslavuj@inf.uniri.hr

ABSTRACT

We present an improved version of our web-based intelligent tutoring system integrated with data mining tools. The purpose of the integration is twofold; a) to power the systems adaptivity based on SPM, and b) to enable teachers (non-experts in data mining) to use data mining techniques on a daily basis and get useful visualizations that provide insights into the learning process/progress of their students.

Keywords

Web based intelligent tutoring system, data visualizations, visual analytics.

1. INTRODUCTION

Our proposed solution to objectives put forth in [5] is the integration of our web-based ITS with standalone data mining tools Weka[3] and SPMF[2]. We developed an integration module that enables continuous communication with the DM tools without implementing any specific algorithm into our application or changing the original DM code. The architecture of the integrated system is displayed in Figure 1. Functionalities that rely on data mining results for students and teachers are marked with asterisks. We will elaborate on these in the next sections. Our web-based intelligent tutoring system (ITS) provides a platform for learning on ill-defined domains [4] i.e. domains that consist of a number of knowledge units (KUs) that do not have a set order in which they have to be learned, but instead the system relies on a domain expert to define the structure of the domain. The learning process is started by selecting a KU to which the system responds by displaying the various types of learning materials created by the teacher. Afterwards, the student proceeds to the assessment module. The system will first ask the student a question about the KU that was learned, followed by an initial question for every KU that is below the current KU in the domain structure created by the teacher. In this way the system checks whether the student understands all the underlying concepts. This list of KUs is currently the same for all students. We aim to make this part dynamic (see Section 3) in order to make the system more adaptive and increase the efficiency of the whole system. If the student offers an incorrect answer to any of the initial questions, he/she is transferred to learning that KU and the whole process is repeated.

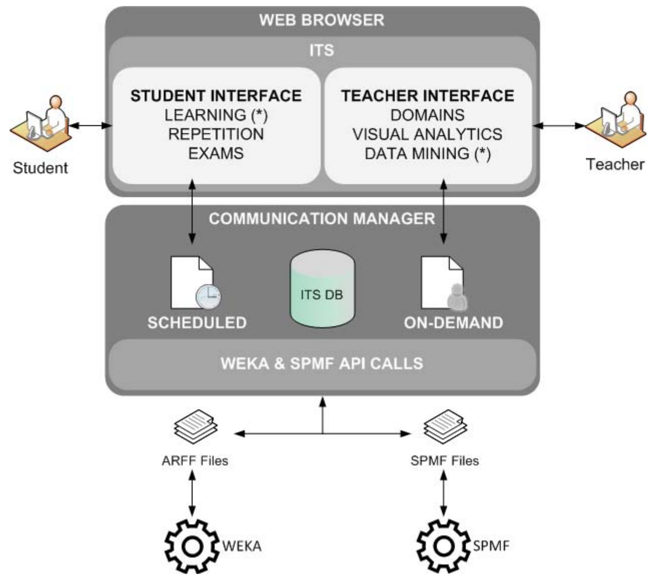

Figure 1. Overall system architecture

No matter how many levels down the hierarchy the student is taken by answering initial questions incorrectly, the system will always return to the starting KU and finish when all the initial questions have been answered. Once the student reaches the KU threshold, the system will stop displaying that KU later in the learning process in order to avoid tediousness and repetition.

2. VISUAL ANALYTICS FOR TEACHERS

At the time of writing the visual analytics section for teachers had a number of visualizations and a clustering section that provide useful insights into the activity of the students and the learning process as a whole. When they start the analytics module, teachers are presented with a compact report containing columns on the number of learning and repetition activities the student performed, number of correct, incorrect and unanswered questions, and the total time spent learning. Each of the columns can be expanded into a sortable, searchable, heat mapped table to get a detailed view about the student's activity. Figure 2 represents the expanded report on the number of learning sessions and repetitions for all the KUs in the domain. Another part of the visual analytics module is the chart section. There are a number of activity charts that can reveal the activity levels of the whole group or individual students (Figure 3).

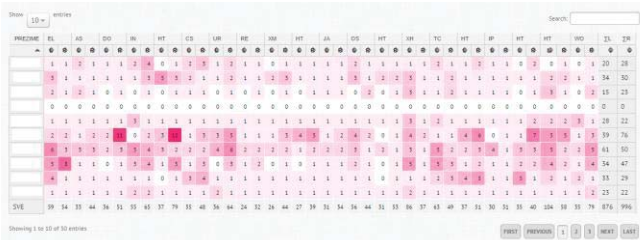

Figure 2. Detailed report on learning (all KUs,all students)

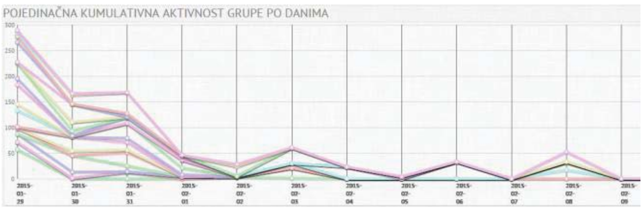

Figure 3. Cumulative group activity by days

The clustering analysis is currently based on a fixed number of features (the ones mentioned in the compact report), but in the next development iteration it will be completely interactive so that the teacher will be able to select features as well as the number of clusters before starting the analysis. When the teacher starts the clustering, the system invokes the communication manager which converts the data to the appropriate file format for either Weka or SMPF, writes the file to the file system and then performs the appropriate API call in the shell command line.

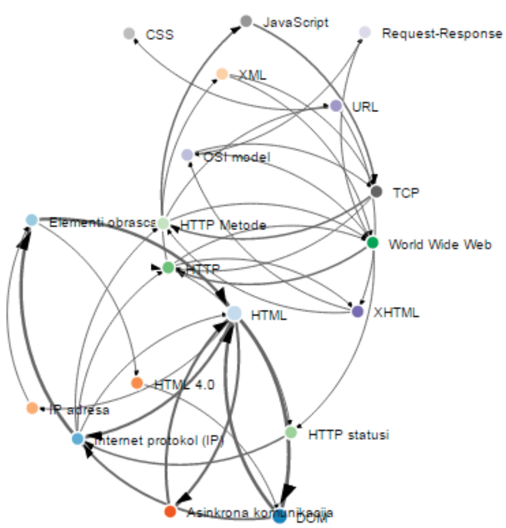

Figure 4. Visualization of student learning paths

The DM tool runs the required algorithm on the data using the sent parameters, and produces the output file. The file is then read, formatted and then returned to the teacher interface where they are displayed as a table with five columns containing cluster names, clusters centroids and students belonging to each cluster. The teachers using the system had no problem identifying inactive students, best students, the average students (largest cluster) and students that were “gaming” the system - students with low number of questions answered and very small amount of time spent learning – they started using the system at the last minute and probably obtained the answers to some questions. This can be confirmed by analyzing the heat maps and activity charts of those students.

3. DM-POWERED PATH OPTIMISATION

The next goal of our research is to create a more adaptive tutoring system in order to: a) increase the quality of learning, b) reduce time needed to acquire the domain knowledge. The set hypothesis is that each student creates a unique path through the structure of KUs. By scheduling a daily analysis of all these paths using SPM algorithms, we can find frequent learning paths. Next, we need to evaluate these paths in order to differentiate between paths that are frequent because a number of students are struggling with a difficult KU without making much progress through the domain from paths that show efficient behaviors that result in significant progress. We are currently developing an algorithm that will perform these evaluations by taking into account a number of learning performance indicators in order to produce a path score. When we get a list of evaluated frequent sequences and students clustered by their activity and effectiveness levels, we can alter the list and order of KUs to be learned in order to help the student follow an optimized path through the knowledge domain. Clustering of students gives us a finer level of granularity so we can offer different modifications to different groups of students. At this moment we run the SPM algorithms to get the frequent patterns and visualize them (Figure 4) using D3JS [1].

4. CONCLUSION

The main advantage of the system is that we can use any of the many SPM and clustering algorithms provided by integrated DM tools. In the future we will complete the SPM based adaptive path optimization component and perform experiments to verify its efficiency.

5. ACKNOWLEDGMENTS

This research is a part of the Project "Enhancing the efficiency of an e-learning system based on data mining", code: 13.13.1.2.02., funded by the University of Rijeka, Croatia.

6. REFERENCES

[1] Bostock, S. M., 2014. D3JS Data Driven Documents. http://d3js.org.

[2] Fournier-Viger, P., et al., 2013. SPMF: Open-Source Data Mining Library. http://www.philippe-fournier-viger.com/spmf/.

[3] Hall, M., et al., 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11, 1.

[4] Lynch, C., et al., 2006. Defining Ill-Defined Domains; A literature survey. In *Proc. Intelligent Tutoring Systems Ill-Defined Domains Workshop*, Taiwan, 1-10.

[5] Romero, C., Ventura, S., 2010., Educational data mining: A review of the state-of-the-art. *Transactions on Systems, Man, and Cybernetics,* , vol. 40, 6, 601-618.