Students Activity Visualization Tool

Marius Ștefan Chirițoiu
University of Craiova
Bvd. Decebal, 107
+40-251 438198
cmariusstefan@gmail.com

Marian Cristian Mihăescu
University of Craiova
Bvd. Decebal, 107
+40-251 438198
mihaescu@software.ucv.ro

Dumitru Dan Burdescu
University of Craiova
Bvd. Decebal, 107
+40-251 438198
burdescu@software.ucv.ro

ABSTRACT

One of the primary concerns in on-line educational environments is the effective and intuitive visualization of the activities performed by students. This paper presents a tool that is mainly designed for the use of professors in order of assist them in a better monitoring of students activity. The tool presents in an intuitive and graphical way the activities performed by students. The graphical presentation creates a mental model of the performed activities from the perspective of former generations of students that followed the same activities. The tool integrates k-means clustering algorithm for grouping students and for facilitating the customization of parameters and number of clusters that are displayed.

Keywords

activity visualization, k-means, e-Learning, PCA.

1. INTRODUCTION

One of the main issues of on-line educational environments is the effective and intuitive visualization of activities performed by students. For an e-Learning platform which has an average of more than 100 students per module it may become quite difficult for the professor to visualize the on-line activity performed by each student at a time or by all students at one time. The paper presents a tool that improves the productivity of a professor by providing an effective way of visualizing and interacting with the students.

This paper presents a tool that displays in a graphical format the activities performed by students. There are several characteristics of the tool that make it user friendly and very efficient in presenting in synthetic form the results. The main characteristic consists of the fact that the display is in 2D (2-dimensional) space and for each coordinate a single parameter is used. The display presents a specified number of groupings according with the settings of the professor. The number of groupings (i.e., clusters) represents the main parameter for the clustering algorithm that effectively places the students into clusters. For each cluster there is clearly presented the centroid. The students from the same cluster are presented with the same distinct geometric sign in color and shape. Each cluster is divided into three areas: center, close area and far area. Each area gathers students that have the same behavioral pattern regarding the activities performed within the on-line educational environment.

2. RELATED WORKS

There are several examples of educational data mining tools. The latest research trends place an important emphasis on developing tools that successfully integrate and prove the latest findings in the domain. In [1] there is presented a section with the more than twenty educational data mining tools. Among them, the ones that are more closely to the tool presented in this paper is GISMO [2], EDM Visualization Tool [3] or SNAPP [4].

GISMO is a tool whose main purpose is to visualize what is happening in distance learning classes. EDM Visualization Tool is mainly designed to visualize the process in which students solve procedural problems in logic. The SNAPP tool may be used to visualize the evolution of participant relationships within discussions forums.

3. SOFTWARE ARCHITECTURE

The application is divided into packages that contain classes that implement related functionalities. The main classes that perform the business logic of the tool are *ClusteringServlet*, *RunScheduledJobServlet*, *BuildArffFileScheduledJob*, *BulidClusterersScheduledJob*, *KMeansClustererStart*, *ClientStervlet* and *ClustererClientApplet*.

The web server administration interface (*index.html*) allows building a number of clusters and viewing them. This is performed by the *ClusteringServlet* which in turn uses the *KMeansClustererStart* class. *KMeansClustererStart* class generates the clusters (i.e., the model) based on ARFF file using KMeans algorithm. It also contains various methods for manipulating clusters data. In this class PCA (Principal Component Analysis) algorithm is used to reduce the dimensionality (number of attributes) of a given dataset.

The *RunScheduledJobsServlet* class is a servlet that starts at server startup and runs the scheduled job at specified time. The time and frequency are specified in a *xml* configuration file. The scheduled jobs are represented by *BuildArffFileScheduledJob* and *BuildClustersScheduledJob* classes. These, as their name implies, deal with building the *arff* file from database and building clusters from *arff*. The *BuildArffFileScheduledJob* class uses *ArffGenerator* class for building the *arff* file containing the training dataset.

On the client side we have a java applet that runs in an Internet browser. It connects to server, takes data needed and displays the students grouped according to certain features chosen by the professor. The interface of the client application allows specifying data for a new student and viewing its position on the chart (in the cluster to which it belongs).
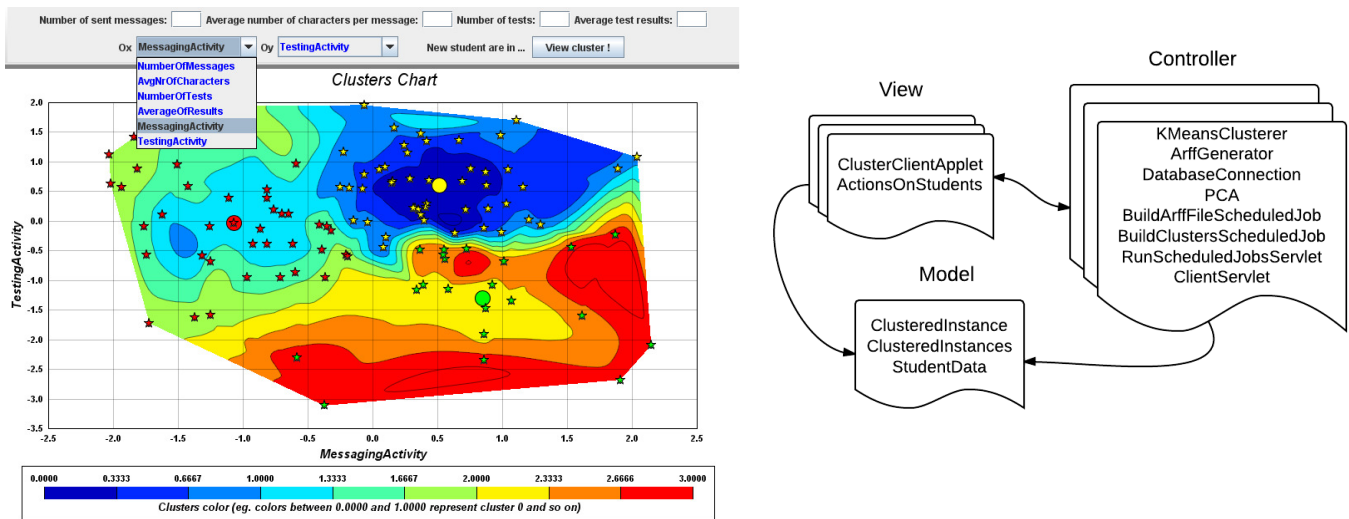
Figure 1. GUI of the visualization tool & Soft Architecture

4. SOFTWARE TOOL

Figure 1 presents the GUI of the visualization tool. In this running there are three clusters of students built according to the two features of the axes (i.e., testing activity and messaging activity). The application allows the professor to select other features by which to build the clusters. When you keep the mouse pointer over a point on the graph you can see the features values for that student. The points representing the students have different colors for each cluster. For a better visualization each cluster is divided into three areas: center, middle area and far area. Each area is colored differently giving thus intuitive information regarding how close from the centroids are the points belonging to a cluster.

The tool uses a total of six features with which we can build clusters as presented in figure 1. The last two features are composed features resulting by combining two simple features using Principal Component Analysis (PCA). The *MessagingActivity* feature is computed as a combination between the *NumberOfMessages* feature and *AvgNrOfCharacters* feature. In the same way, the *TestingActivity* feature is computed as a combination between *NumberOfTests* and *AverageOfResults* feature.

If features values are provided for a new student the tool places a big X mark with the same color as other instances from the same cluster in corresponding position and thus the cluster is immediately determined. After viewing clustered students, the teacher can select one or more students from the chart and save their data in a PDF or send an e-mail to them.

The tool may be used for two purposes. One regards easy visualization of the student's activity based on different criteria. Once the visual information is retrieved, the tool may be used to interact (i.e., send messages) with a specific set of students that may be easily selected. The tool may be also successfully used for outlier detection, which in our case is represented by students that hardly can be assigned to a cluster.

5. CONCLUSIONS

This paper presents a visualization tool based on a clustering algorithm. The tool presents the clusters of students in a very intuitive way. The clustered students may be selected and a set of specific actions may be performed: sending messages, export to pdf, etc. In future, the tool may be extended by integrating other features for data representation and by providing other advanced functionalities for professors.

6. REFERENCES

[1] Cristobal Romero and Sebastian Ventura, Data mining in education, WIREs Data Mining Knowl Discov 2013, 3: 12–27 doi: 10.1002/widm.1075.

[2] Mazza R, Milani C. GISMO: a graphical interactive student monitoring tool for course management systems. In: International Conference on Technology Enhanced Learning. Milan, Italy; 2004, 1–8.

[3] Johnson M, Barnes T. EDM visualization tool: watching students learn. In: Third International Conference on Educational Data Mining. Pittsburgh, PA; 2010, 297–298.

[4] Bakharia A, Dawson S. SNAPP: a bird's-eye view of temporal participant interaction. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge. Vancouver, British Columbia, Canada; 2011, 168–173.