

Players Performances Analysis based on Educational Data Mining Case of Study: Interactive Waste Sorting Serious Game

Elaachak Lotfi
Computer Science, Systems
and Telecommunication
Laboratory (LiST)
Faculty of Sciences and
Technologies
Abdelmalek Essaadi University
Tangier, Morocco

Belahbib Amine
Computer Science, Systems
and Telecommunication
Laboratory (LiST)
Faculty of Sciences and
Technologies
Abdelmalek Essaadi University
Tangier, Morocco

Bouhorma Mohammed
Computer Science, Systems
and Telecommunication
Laboratory (LiST)
Faculty of Sciences and
Technologies
Abdelmalek Essaadi University
Tangier, Morocco

ABSTRACT

Serious games have become one of the powerful tools in the education field, view of their capability to transmit the knowledge to the players/ students, but to judge if a given serious game is effective there must be a system that analyzes the performances and behaviors of the players, to see their level of understanding about a particular topic proposed by the serious game. In this perspective of research and development this paper presents a method for analysis concerning the performances of serious game players, based on educational data mining, with the aim of helping the instructors and the experts to improve their strategies of teaching. An evaluation of how our method proved successful with an outlook on future research concludes this paper.

General Terms

Player Performances in Serious Games

Keywords

Serious Games, Waste Sorting, Educational Data Mining, K-means, Assessment, Player Performances

1. INTRODUCTION

The serious games are designed to have an impact on the target audience, which is beyond the pure entertainment aspect [1, 2]. This kind of video games, have for mission the transfer of the knowledge to the learners in an entertaining way. Among the features proposed by serious games there are : the attractiveness, the interactivity, the system of assessment, and the transfer of different pedagogical messages in different way, with those features serious games have become one of the powerful tools in educational field this last decade.

As known the most important criteria that judges the efficiency of a serious game is its potential to teach new skills and its capability to transmit the knowledge to the learners in a playful way, in addition, among the challenges that faces instructors and experts is how to evaluate the learning outcomes in order to identify if the given serious game is suited for a given goal or field, for this reason many studies have been done concerning the potential educational benefits that serious games may offer, but the problems are finding reliable measures for learning, and one of the biggest challenges has been finding accurate and reliable measures for fun and learning [3]. In addition the method followed to analyze such measurement is can be decisive view of its importance and its role that plays for helping the instructors

and experts to improve their methodology of teaching according to the performances and the behaviors of the students / players.

In summary, this paper presents a method that analysis the player performances based on educational data mining of a serious game developed by our research team and dedicated to children, the proposed serious game is concerning the protection of environment filed, and its concept is simple, the player have to drag and drop different objects and put them in the correct container according to their types, but the novelty is the way of interaction between the player and the serious games by using a controller that senses how naturally move player hands and fingers to drag and drop different objects, all gestures and behaviors of the players will be saved in the database then processed by an algorithm called k-means for clustering, as a final step the result given by k-means clustering algorithm will be used by the instructors and experts to classify their students according to their performances, in order to detect problems met by the learners/ players during their learning process.

2. INTERACTIVE WASTE SORTING SERIOUS GAME

Waste sorting has and should become part of our daily life to improve our living environment. With the importance of the waste sorting and the benefits that it presents, to this effect many instructional experts have found that teaching the basics of waste sorting for kids dice their young age can be beneficial for the environment and the economy. For this reason there is no more robust way to learn the basics of waste sorting better than the video games, view to their advantages such interactivity and playability that attracts the intention and desire into the players to play more. In this perspective of development our research team has developed an interactive web based serious game for waste sorting dedicated for children that will be described in this section.

2.1 The Concept of Waste Sorting Serious Game

The main objective of the proposed serious game Fig.1 is to teach kids about recycling different waste. The player should sort different waste into trash, paper, plastic, metal, glass, and organic, etc. The sorting is done by catching different objects generated randomly and dropping in the appropriate container according to their types, this mechanism will be done by using a tool called a leap motion controller.

The waste sorting serious game will be equipped by the timer, and the assessment system that evaluates the players according to their performances if they make a good choice the reward will be the gain of some points, although if the opposite case the punishment will be the loss of some points. With the assessment system, the timer, and the interactivity based on hand movement the proposed serious game will be more challenging and attractive spicily for kids, it will allow them to live a beneficial and unforgettable experience.

The proposed serious game has been developed by JavaScript API, therefore, it need just a web browser to be run.



Fig. 1: Screen shoots from waste sorting video game.

2.2 Interactivity with Leap Motion Controller

The Leap Motion controller is a small device that can be connected to a computer using a USB. It uses infrared (IR) imaging to determine the position of predefined objects in a limited space in real time. It can then sense hand and finger movements in the air above it, and these movements are recognized and translated into actions for the computer to perform.

According to the official information founded in the official web site of leap motion [4], the Leap software analyzes the objects observed in the device's field of view. It recognizes hands, fingers, and tools, reporting discrete positions, gestures, and motion. The controller's field of view is an inverted pyramid centered on the device Fig. 2. The effective range of the controller extends from approximately 25 to 600 millimeters above the device. The controller itself is accessed and programmed through Application Programming Interfaces (APIs), with support for a variety of programming languages, ranging from C++ to Python and JavaScript. The positions of the recognized objects are acquired through these APIs. The Cartesian and spherical coordinate systems used to describe positions in the controller's sensory space.

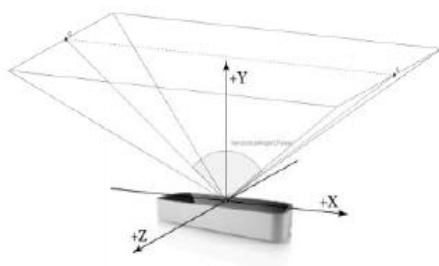


Fig. 2: Leap Motion's Field Of View

With the features that leap motion controller offers and with the use JavaScript application programming interface, we have integrated it in the proposed video game, therefore, the player has to move his hand and catch the random generated objects in order to drag and drop them, then place them in a correct container. With this possibility the proposed video game will become more interactive and so close to the real case. That will create envy into the player to play more. In addition, this concept will allow us to save all gestures done by the players during a sequence of video game; this data will be used by educational data maiming to understand the player's behaviors and also to analyze their performances, in the next section we will detail more the process of educational data mining.

3. EDUCATIONAL DATA MINING

The Data Mining is the process of analyzing data from different perspectives and summarizing the results as useful information. It has been defined as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [5, 6].

Among the most famous branches of data mining there are the educational data mining "EDM" that describes a research field concerned with the application of data mining, machine learning algorithms and statistics tools to information generated from educational area.

Other definition of educational data mining as tool of Mining in educational environment, concern with developing new methods to discover knowledge from educational databases [7].

It's an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [8].

3.1 The Knowledge Discovering Process

The knowledge discovering process KDD includes selecting the data used in data mining process, this data can be obtained or extracted from different and heterogeneous data sources e.g. "database, xml files and other resources". As shown in Fig.3 the data mining is an essential step in the process of knowledge discovers.

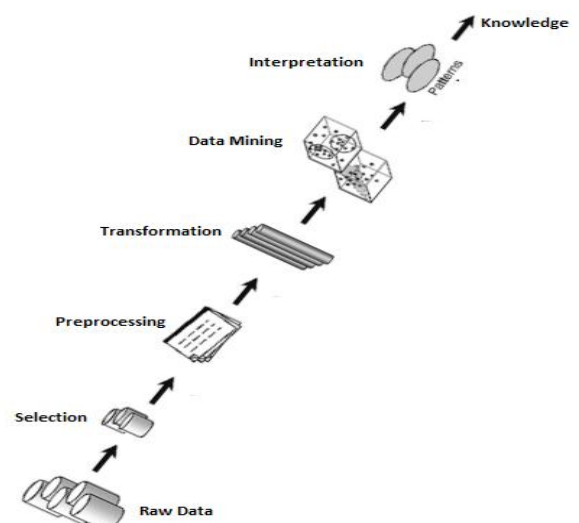


Fig.3: knowledge discovering process Data mining method

3.2 Data Mining Techniques

As known the main objective of data mining including educational data mining is extraction of data to predict unknown or future values of the attributes, and also to describe the data in a manner understandable and interpretable to users. For this reason there are several methods regrouped in several different classes described below: Classification, Clustering, Association rule, Regression, and others machine learning algorithms like Neural networks, BayesNet, etc.

Association rule

Mining association rules searches for interesting relationship among items in a given data set [9]. It's used to find frequent item set finding among large data set. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little value.

Types of association rule:

- Multilevel association rule.
- Multidimensional association rule.
- Quantitative association rule.

Clustering

Clustering is applied to position items of heterogeneous data resources in to specific groups according to some attributes. By using clustering techniques it's possible to identify dense and sparse regions in object space, and correlations among data attributes. It can be used as tool to distinct groups or classes of object but it becomes costly to clustering.

Types of clustering methods:

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

Classification

Classification is a data mining task that predicts group membership for data instances [10]. It's the most commonly applied data mining technique, which employs a set of pre-classified pattern to develop a model that can classify data.

Types of classification models

- Decision tree
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

Regression

Regressions techniques can be adapted for prediction. In general the regression analysis can be used to model the relationship between independent and dependent variables. In data mining independent variables are attributes already known and response variable are what users want predict [15].

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

Neural networks

The typical neural network consists of nodes that are connected to each other and exist in several different layers, resulting in it being often referred to as a Multi Layered Perceptron (MLP) network. These layers are the input layer, the hidden layer, and the output layer. Each of these layers has a design specific amount of individual nodes in them. An individual node works much like its biological counterpart the neuron. It receives input from a multitude of different weighted input connections, sums these inputs and then produces an output that serves as input for other nodes. This output is generally normalized to be between -1 and 1 and typically a sigmoid function of the type discussed in can be used for this [11].

Types of neural networks

- Back Propagation

4. METHOD

As mentioned in the section above the knowledge discovering process data mining method is composed of different steps, in this section we will detail the process followed in order to analyze the players performances by using educational data mining especially k-means algorithm for clustering.

4.1 The Knowledge Discovering Process

As mentioned above the serious game will be equipped by a database, see the class diagram in Fig.4, the information like gestures, score, number of good choices, bad choices, and all the behaviors of the player, in addition to the personal information as age and name will be saved in this database, all of this information will help instructors to analyze player performances by feeding the k-means algorithm in order to group the students according to their performances.

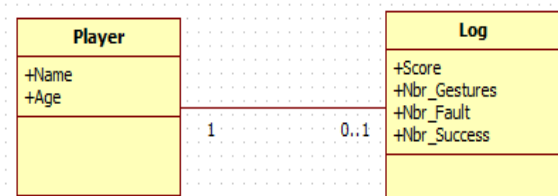


Fig.4: Class diagram of the log system of the proposed serious game

4.2 WEKA Data Mining Software

The Waikato Environment for Knowledge Analysis “WEKA” came about through the perceived need for a unified workbench that would allow researchers easy access to state of the art techniques in machine learning. It is recognized as a landmark system in data mining and machine learning [13]. It has achieved widespread acceptance within academia and business circles, and has become a widely used tool for data mining research. The book [14] that accompanies it is a popular textbook for data mining and is frequently cited in machine learning publications. This Workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.

The Graphical user interface Chooser WEKA’s graphical start point has undergone a redesign and now provides access to various supporting user interfaces, system information and logging information, as well as the main applications in WEKA. Fig.5 shows the revamped GUI Chooser.

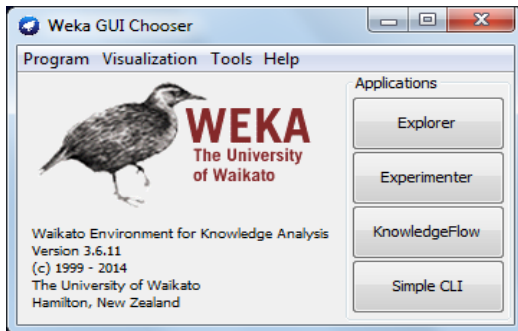


Fig.5: The GUI Chooser

Scatter plots, ROC curves, trees and graphs can all be accessed from entries under the “Visualization” menu. The “Tools” menu provides two new supporting GUIs: SQL viewer and Bayes network editor. In addition of other features that makes the use of WEKA more helpful for the users.

4.3 Definition of K-Means Clustering

The k-means algorithm [12] selects randomly k number of objects, each of which initially represents a cluster mean or center, an object is assigned to the cluster to which it is most similar, based on the distance between the object and cluster mean. Then it computes new mean for each cluster. This process iterates until the criterion function converges, the flowchart of k-means algorithm shown in Fig. 6.

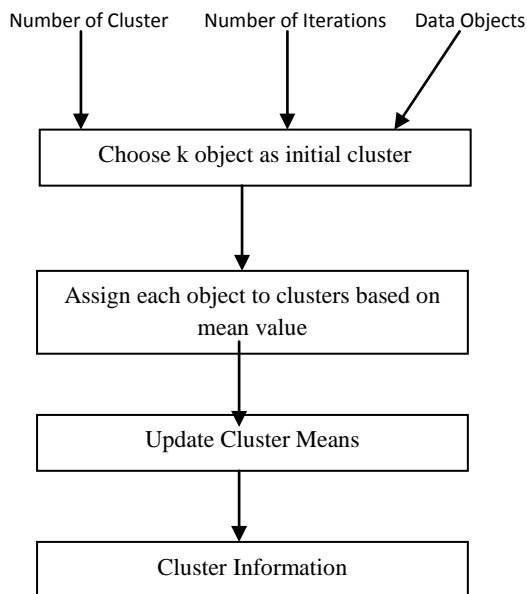


Fig.6: Flowchart of k-mean algorithm

4.4 Clustering Players with the use of k-means Algorithm

As mentioned above the information saved during the serious game sequence from each player, are: score, number of gestures, number of good choices, number of bad choices and the age of the player. All this information will help teacher to analyze the performances of the players, the method consist to group the players according to their performances.

The choice of k-means algorithm is due to the fact that all information recorded are digital, because some implementations of K-means only allow numerical values for attributes.

With the use of the WEKA graphical user interface, the instructor can load data from database as shown in Fig. 7.

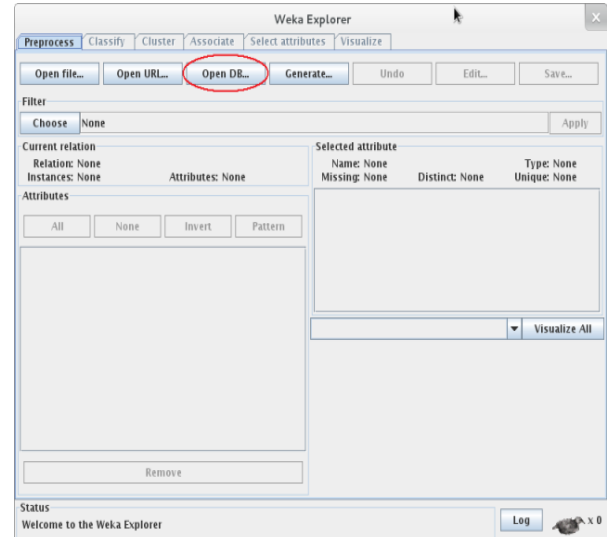


Fig.7: WEKA Explorer interface with the database loaded

After the load of the data from the database, the user can shows the attribute that will be used by k-means algorithm in Fig. 8.

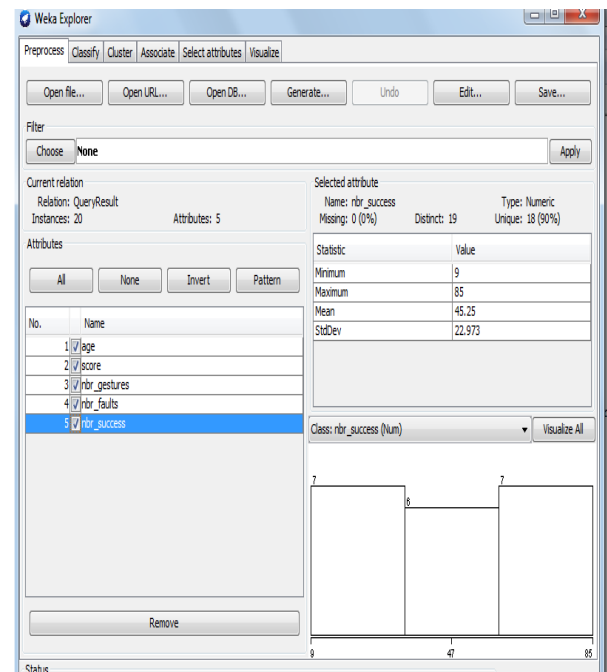


Fig.8: WEKA Explorer Preprocess Interface

The Cluster interface from the WEKA GUI proposes several algorithms among them there are: SimpleKmeans, EM, Cobweb, xMeans, etc. In our case the user chooses the SimpleKMeans algorithm, and in addition he can pass the parameters like number of clusters, seed and number of iterations, etc. Fig. 9.

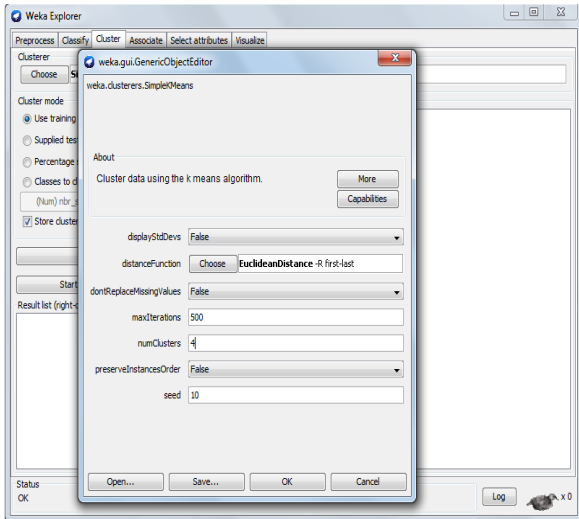


Fig.9: WEKA Explorer Cluster Interface

The simpleKMeans algorithm automatically handles a mixture of categorical and numerical attributes. Furthermore, the algorithm automatically normalizes numerical attributes when doing distance computations. The WEKA simpleKMeans algorithm uses Euclidean distance measure to compute distances between instances and clusters. In the next section we will detail and discuss the result obtained by the chosen algorithm.

5. RESULTS

Once the options have been specified as number of clusters “4 clusters”, seed and the “Use training set” number is selected. The clustering algorithm will be ran, the result of the algorithm shown in Fig 10.

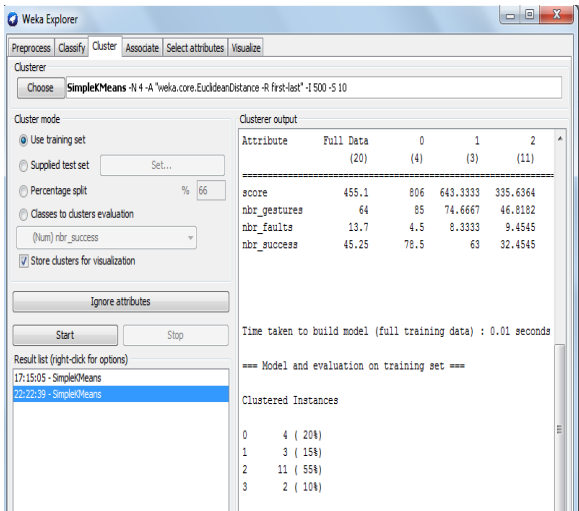


Fig.10: Result of K-MEANS Clustering

There are 20 players concerned by this study, as generated result there are four clusters, according to the score, number of gestures, number of good choices and bad choices. The clustering produced by k-means shows 20% (4 instances) in cluster 0, 15% (3 instances) in cluster 1, 55% (11 instances) in cluster 2 and 10% (2 instances) in cluster 3. In our case the simple k-means algorithm has done 7 iterations to cluster results, concerning the within cluster sum of squared errors value is 1.1401, and the time taken to build model is about 0.01 seconds, the information about Cluster centroids was detailed in Table1.

Table 1. Cluster centroids

Attribute	FD	0	1	2	3
	(20)	(4)	(3)	(11)	(2)
Score	455.1	806	643.3	335.6	128
Nbr_gestures	64	85	74.6	46.8	100.5
Nbr_faults	13.7	4.5	8.3	9.4	63.5
Nbr_success	45.2	78.5	63	32.4	22.5

For more details the result shows that in cluster 0 the score is 805, the number of gestures is about 85, number of good choices is 78.5, and number of bad choices is 4.5. In cluster 1 the score is 643.33, the number of gestures is about 74.66, number of good choices is 63, and number of bad choices is 8.33. In cluster 2 the score is 335.6, the number of gestures is about 46.8, number of good choices is 32.45, and number of bad choices is 9.45. In cluster 3 the score is 128, the number of gestures is about 100.5, number of good choices is 63.5, and number of bad choices is 22.5. The cluster 0 covers 20% of players cluster 1 covers 15% of players, cluster 2 covers 55% of players and cluster 3 covers 10% of players.

Interpreting the results given by the simpleKMeans algorithm, the cluster 0 represents the good players/students who understood all the basics of waste sorting, view to their score and their performances. For the players that belong to the cluster 1, are generally good, but they need some guidance and explanation concerning the basics of the waste sorting. Taking the case of cluster 2 the students that belong to that cluster, have an average level, they have few problems regarding their understanding about the proposed topic and they need an explanation and assistance to understand the basics of waste sorting. For the final cluster 3 the most of the students that belong to this cluster have several problems and difficulties, they have chosen the objects randomly and without thinking. They need explication on basics of waste sorting, and in addition of a special assistance, in order to increase their level of comprehension. The percentage of each cluster is shown in graph below Fig 11.

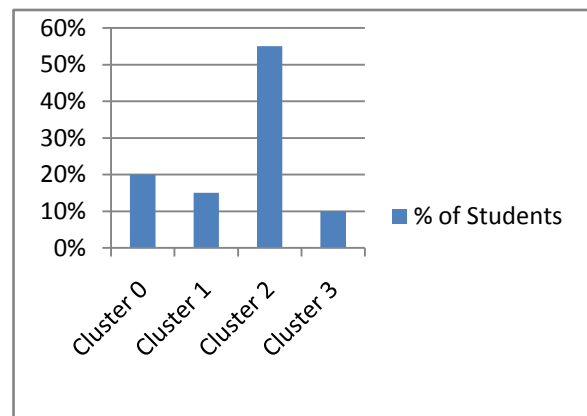


Fig.11: Graph concerning percentage of students in each cluster

Another way of understanding the characteristics of each cluster is through visualization the Fig.12 shows cluster representation according to the score compared to the number of gestures, we can visualize other result according to different combinations of score, number of gestures, and number of good and bad choices.

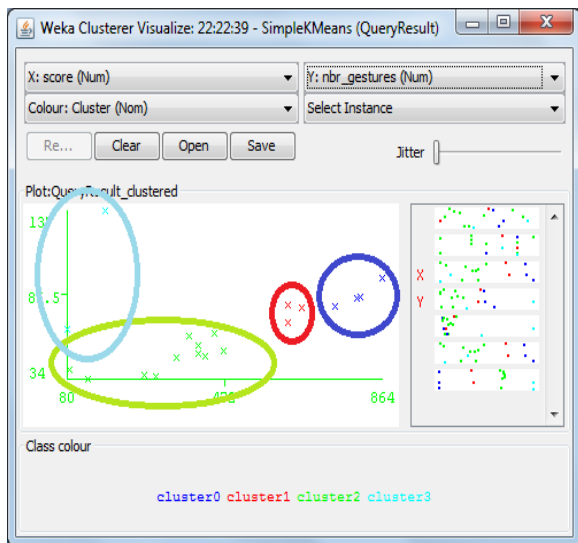


Fig.12: Visual result

The instructor can add age in his analysis to get information that will help him change his course to be better suited for these students. This analysis method based on collecting information about students during a sequence of serious game with the use of the educational data mining can be operated as needed and according to the learning strategy of the instructors.

6. CONCLUSIONS

To conclude this paper, the use of the educational data mining applied on serious games, can be beneficial for the instructors and the experts, and as more and more data is collected from database, the instructor can analyze the behaviors and performances of their students/players to understand their ways of learning, and have a global view of their interaction with the serious game, this method will allow the instructors to improve their own strategy of teaching.

Concerning the perspectives we will envisage to equip the proposed serious game by other algorithms of educational data mining, e.g. “classification, association rules” combined with learning analytics techniques in order to have other information about players outcomes, in addition, the proposed serious game will be equipped by an inference engine that will analyze players behaviors and gives them several indication according to their performances, this new approach will be a shift towards smart serious games.

7. ACKNOWLEDGMENTS

This research paper is made possible through the help and support from the students of computer engineering. We gratefully acknowledge the support of the bachelor, master students and all other participants.

8. REFERENCES

[1] J. P. Gee, What Video Games Have to Teach Us about Learning and Literacy, Palgrave MacMillan, New York, NY, USA, 2007.

[2] Horton, M., Read, J. C. & Sim, G. 2011. Making your mindup?: The reliability of children's survey responses. In Proceedings of the 25th BCS Conference on Human Computer Interaction (BCS-HCI '11). British Computer Society, Swinton, UK, 437-438.

[3] Kremer, K. 2012. Conducting Game user experience research with preschools, in Proceedings of the CHI Workshop on Game User Research (CHI-GUR 2012), Austin, TX, USA.

[4] Leap Motion Controller. Available online: <https://www.leapmotion.com> (accessed on 21 October 2014).

[5] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthuramy, R. (1996). Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press.

[6] Frawley W., Piatetski-Shapiro G. and Matheus C. (1992). Knowledge discovery in databases : an overview. AI Magazine. 14(3). 57-70.

[7] Erdogan & Timor. (2005). A Data Mining Application in A Student Database. Journal of Aeronautic and Space Technologies July 2005 Vol. 2 Number 2 (53-57).

[8] Margaret H. Dunham “Data Mining Introductory and Advanced Topics”.

[9] Agrawal, R. Imielinski, T. Swami, A., “ Mining Association Rules Between Sets of Items In Large Databse”. In proceedings of the ACM SIGMID Conferences on Management of Data, Page 207-206 Washington DC. May. 1993.

[10] Han, j. And Kamber, M. “Data Mining: Concepts and Techniques” , 2end Edition. The Morgan Kaufman Series in Data Management Systems, Jim Gray, Series Editor. 2006.

[11] Lee, H. M, Huang, T.C., and Chen, C.M., “Learning Efficiency Improvement of Back Propagation Algorithm by Error Saturation Prevention Method”. International Joint Conference on Neural Networks 1999 (IJCNN '99), Volume 3, 10-16 July 1999, p. 1737 – 1742.

[12] Alsabti, K., Ranka, S., & Singh, V. (1998). An efficient k-means clustering algorithm. IPPS/SPDP Workshop on High Performance Data Mining. IEEE Computer Society Press.

[13] G. Piatetsky-Shapiro. KD nuggets news on SIGKDD service award. <http://www.kdnuggets.com/news/2005/n13/2i.html>, 2005.

[14] I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2 edition, 2005.

[15] Vivek Bhambri , Application of Data Mining in Banking Sector, Dept. of Computer Sciences, Desh Bhagat Institute of Management and Computer Sciences, Mandi Gobindgarh, Punjab, India, IJCST Vol. 2, ISSue 2, June 2011.