Student's Academic Performance Analysis using SOM

A. Shaik Abdul Khadir[1] K. Mohamed Amanullah[2] P. Gowri Shankar[3]
[1,2]Associate Professor [3]M.Sc M.Phil
[1,2,3]Department of Computer Science & Engineering
[1,2,3]Bishop Heber College (Autonomous)

Abstract— The ability to monitor the progress of students' academic performance is a critical issue and monitoring student academic performance is not an easy task. Many factors could act as barriers to students attaining and maintaining high scores throughout the academic career. Moreover such type of problems could be solved with the help of data mining techniques. The prediction of academic results shows the students overall academic performance during their tenure and enables them to cope with academics easily. The proposed study aims to develop an effective decision classifier for monitoring and predicting students' academic performance using clustering techniques. One such method investigated in this model is Self Organizing Map or SOM clustering technique. This method is applied to the student data set and the results are predicted.

Key words: Data Mining, SOM, Self Organizing Map, Students Academic Performance

I. INTRODUCTION

Student academic performance is an important factor for gaining employment apart from the fact of better understanding of concepts and learning. In order to improve and better gauge the students' performance academic professionals use various techniques and perform assessment measures. These factors could be targeted by the faculty members and help to develop strategies to improve student learning and also improve their academic performance by way of monitoring the progression of their performance. The performance evaluation is one of the bases to monitor the progression of student performance. Grouping of students into different categories according to their performance has become a complicated task. Data mining has been used as an accurate analysis tool to test and gauge such performances. This can be achieved by applying data mining techniques the academic performance can be analyzed from different angles.

Normally it is difficult to obtain a comprehensive view of the state of the students' performance. Also simultaneously discover important details from their time to time performance. Use of cluster analysis to segment students in to groups according to their characteristics and use decision tree for making meaningful decision for the student's. With the help of data mining methods, like SOM clustering algorithm it is possible to discover the key characteristics from the students' performance and possibly use those characteristics for future prediction.

Cluster analysis could be divided into hierarchical clustering and non-hierarchical clustering techniques. Some of the commonly used hierarchical techniques are single, complete, average, median, and ward etc. Examples of Non-hierarchical techniques include SOM or Self Organizing maps, simple k-means clustering, and adaptive k-means clustering, fuzzy logic clustering and other clustering techniques. To determine which algorithm is suitable for our purposes like academic analysis depends on the function type and that of the dataset available along with the particular purpose of analysis. Further the stability of the clusters can be investigated in simulation studies. The problem of selecting the "best" or accurate algorithm for clustering the dataset is not easy is purportedly a difficult one. For the purposes considered like student academic performances SOM (Self Organising Map) appears to be a good clustering algorithm and ideally should produce clusters with distinct maps with boundaries ,which should suffice for the purpose.

This analysis of student academic performance could be a consequence of an impressive performance in the coursework exams by a large proportion of students resulting in less variation in the recorded grades. However most importantly learning preferences were found to be independent of both the age and gender of students. The study shows that more the determinants of academic performance need to be investigated and that the coursework grades, though literarily insignificant were consistently lower than that of the other students.

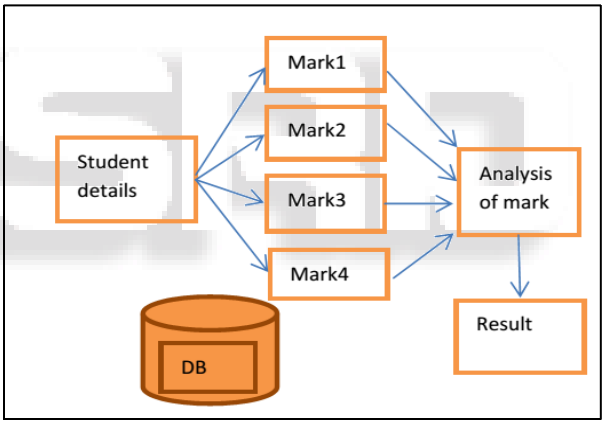

Fig. 1:

II. METHODOLOGY

First the dataset of student performance is prepared with the necessary attributes. This dataset is loaded into the system. The attributes are read and network of clusters are formed. A self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of artificial neural network (ANN) that can be trained on a data set by using unsupervised learning to produce a low-dimensional graphical representation of the input attributes using distances of the training samples which is represented on a map. Self-organizing maps are slightly different from other data mining or artificial neural networks because they use a quick neighborhood function to preserve the coordinate or topological properties of the attributes input in the system. These type of networks learn to form their own classes from the training dataset without taking any external help. In order to achieve this one has to assume that the class membership is defined broadly by the input patterns which share the common features, and enable the network to be able to identify those features across the input range of attributes whose patterns are found or matched. The most interesting attribute of the system is that this is based on a type of learning where the output nodes or points compete amongst themselves in order to be taken as a node on the map and that the node is activated at any point of time. This coordinate point works on the winner-takes all policy and competition can be implemented by having lateral connections or other paths between the attributes or rules. The result is that the map is forced to organise the coordinate points by itself with the given attributes and so or obvious reasons the Self Organizing Map (SOM) is formed.

III. Setting up a Self-Organizing Map

The principal goal of an SOM is to transform an incoming signal pattern of arbitrary dimension into a one or two dimensional discrete map, after which it performs the transformation adaptively to a topologically ordered fashion in the map. Next set up Self Organising Map neurons got from the attributes. This is done by placing the attribute neuron at the nodes of a single or two dimensional lattice structure and move up the order. Further higher dimensional maps are also possible. The attributes in the dataset are tuned to various input patterns called as classes of input patterns during the machine learning or training data set phase. The locations of the attribute neurons so learned or trained (i.e. the selected ones) become ordered as a structure. Thus a meaningful coordinate system for the graph is formed for the selected input features from the dataset which is created on the output screen. The Self Organising Map thus formed shows the student datasets topographic map of the input pattern. The particular kind of Self Organising Map is known as the Kohonen Network, which Map's structure has a feed-forward layer with a single computational layer arranged in rows and columns where each and every neuron is fully connected to all the source nodes in the input layer: So a single dimensional map will just have a single row or a single column in the computational layer compared to other dimensions.

IV. Student Dataset

Students' GPA data collected from the institutes was in continuous form which has to be transformed in to discrete categories. After categorizing the data, it was checked for missing data and it was found that there were no missing attributes. The attributes contained atleast one of the variables factor: however other background data was not considered in the data analysis. The main subjects were considered and the dataset data was further analyzed by selecting relevant items defining a particular factor.

The information gained, with respect to a set of examples is the expected reduction in entropy that results from splitting a set of examples using the values of that attribute. This measure is used for identifying those attributes that have the greatest influence on classification.

The figure shown below the input attributes of the student dataset, whose weighted averages have been taken into account. The next figure shows the graphical map of the attribute neurons considered.

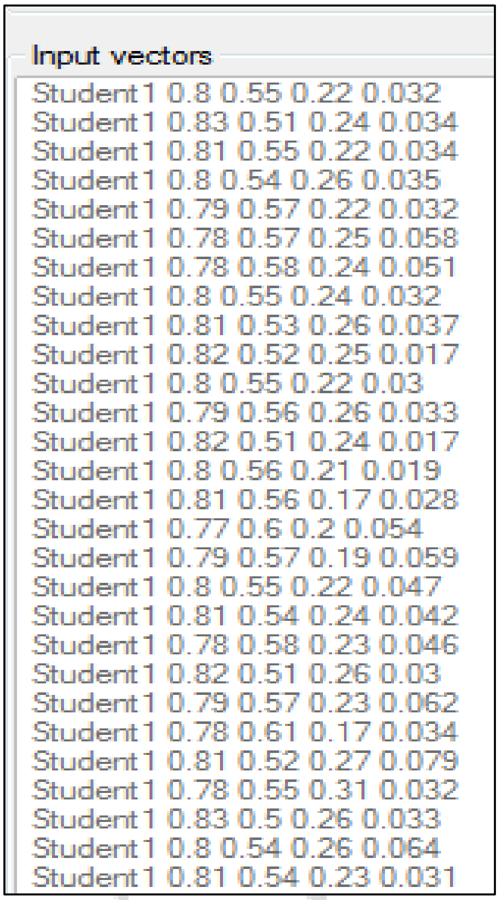

Fig. 2:

Finally the graph of the SOM map of the student dataset is shown in the output alongwith the legend.

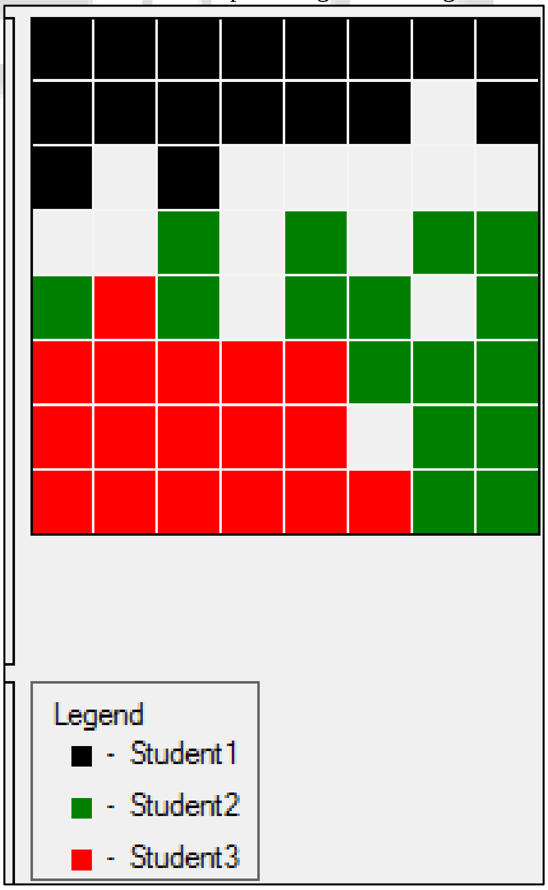

Fig. 3:

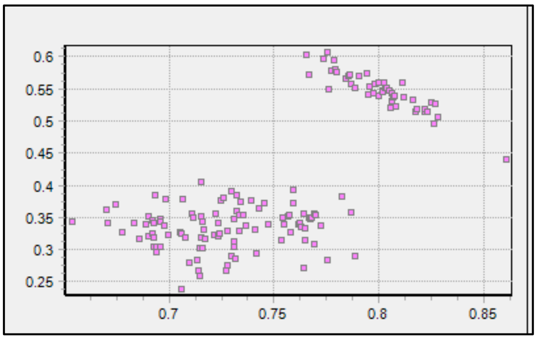

Fig. 4:

V. ALGORITHM STEPS

1) Step 1: First input the number of clusters to group data into and the dataset to cluster as input values
2) Step 2: Initialize the first N clusters
   - Take first n instances of the attributes and then take random attributes of n
3) Step 3: Next calculate the arithmetic means of each cluster formed in the dataset.
4) Step 4: The Self Organising Map then assigns each record in the dataset to only one of the initial clusters

– The next thing is that each record is assigned to the nearest cluster node by using a measure of the distance like Euclidean distance.

5) Step 5: The Self Organising Map model then places each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.

VI. CONCLUSION

Therefore, this proposed SOM clustering algorithm serves as a good benchmark to monitor the progression of students' performance and also enhances the decision making by academic persons to monitor the candidates' performance semester by semester by improving on the future academic results in the subsequence academic session.

REFERENCES

[1] S. Sujit Sansgiry, M. Bhosle, and K. Sail, "Factors that affect academic performance among pharmacy students," American Journal of Pharmaceutical Education, 2006.
[2] Susmita Datta and Somnath Datta, "Comparisons and validation of statistical clustering techniques for microarray gene expression data," Bioinformatics, vol. 19, pp.459–466, 2003.
[3] Rousseeuw P. J, "A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational Appl Math, vol 20, pp. 53–65, 1987.
[4] Sharmir R. and Sharan R., "Algorithmic approaches to clustering gene expression data," In current Topics in Computational Molecular Biology MIT Press; pp. 53-65, 2002.
[5] Mucha H. J., "Adaptive cluster analysis, classification and multivarite graphics,"Weirstrass Institute for Applied Analysis and Stochastics,1992.
[6] Fahim A. M., Salem A. M., Torkey F. A. and Ramadan M. A., "An efficient enhanced k-means clustering algorithm," Journal of Zhejiang University Science A., pp. 1626–1633, 2006
[7] J. O. Omolehin, J. O. Oyelade, O. O. Ojeniyi and K. Rauf, "Application of Fuzzy logic in decision making on students' academic performance," Bulletin of Pure and Applied Sciences, vol. 24E(2), pp. 281-187, 2005.
[8] J. O. Omolehin, A. O. Enikuomehin, R. G. Jimoh and K. Rauf, "Profile of conjugate gradient method algorithm on the performance appraisal for a fuzzy system," African Journal of Mathematics and Computer Science Research," vol. 2(3), pp. 030-037, 2009.

