Prediction of Student’s Academic Performance Based on Adaptive Neuro-Fuzzy Inference

Altyeb Altaher and Omar BaRukab

Department of Information Technology, Faculty of Computing and Information Technology-Rabigh
King Abdulaziz University, P.O. Box 344
Rabigh, 21911, Saudi Arabia

Summary
Prediction of student’s performance is potentially important for educational institutions to assist the students in improving their academic performance, and deliver high quality education. Developing an accurate student’s performance prediction model is a challenging task. This paper employs the Adaptive Neuro-Fuzzy Inference system (ANFIS) for student academic performance prediction to help students improve their academic achievements. The proposed approach consists of two steps. First, results of the students in the previous exams are preprocessed by normalizing their results in order of improving the accuracy and efficiency of the prediction. Second, The ANFIS is applied to predict the students' expected performance in the next semester. Three ANFIS models viz. ANFIS-GaussMF, ANFIS-TriMF and ANFIS-GbellMF that utilized various membership functions to generate accurate fuzzy rules for the prediction process of the student’s performance are used in this research work. The experimental results showed that ANFIS-GbellMF model surpassed the other ANFIS models with a Root Mean Square Error (RMSE) as low as 0.193.

Key words: Adaptive Neural Fuzzy Inference System; Student's performance prediction; educational data mining.

1. Introduction
Student performance prediction is an important factor to deliver high quality education. Educational Institutes are endeavoring to incorporate student Performance prediction into their educational processes for better students' support, by arranging additional efforts for the students with lower performance. Data mining methods are used extensively in many organizations as a promising way for exploring the enormous amounts of data, finding beneficial information to help in making potential decisions. Data mining methods can be applied in educational institutes to advance the students’ performance and the quality of education. Educational data mining which has been used widely in educational area, concern with finding methods for exploring the continuously increasing amount of data in the educational domain to extract significant information and new knowledge to guide the students’ learning and improve the process of learning [1,2,3].

The continuously increasing amount of educational databases, contain potential hidden information that are yet to be discovered to improve students' academic performance. Education data mining explores massive educational databases to extract the hidden significant information for further processing. The smeared information will support many educational processes such as students' performance prediction [6], which enable the educators to explore potential knowledge about students. The classification, clustering, visualization and regression, are various techniques that have been used for extracting the hidden information in educational databases.

Classification has been considered frequently as an important technique in educational data mining. Classification is a method that assigns data values to different predefined Classes, it aims to predict the correct class for each data element. There are several approaches for data classification including artificial neural network, Bayesian classifiers and support vector machine techniques [9]. These approaches have been used widely in educational systems [10–12]. The Neuro-fuzzy inference system (NFIS) is a soft computing tool which combines the fuzzy logic reasoning with the neural network capability of learning, thus the neuro fuzzy inference system handle the disadvantages of both neural networks and fuzzy systems when they are used separately [13]. NFIS has been adapted successfully in applications such as control and classification [14,15].

In this paper, ANFIS is used for students' academic performance prediction in order to help students to improve their achievements. The proposed ANFIS based approach intelligently integrates the capability reasoning of the fuzzy logic with the learning abilities of the neural networks. The main contributions of this paper is the utilization of ANFIS for accurate student performance prediction. Section 2 presents the literature review while the proposed approach is explained in sections 3. Results and discussion are shown in section 4. Conclusion is presented in section 5.

Manuscript received January 5, 2017
Manuscript revised January 20, 2017
2. Literature Review

The rapid advances in education technologies assist the educational institutes to practice the student performance prediction models [4,5]. Several techniques were used to predict students' performance. In [7], Barber and Sharkey used the logistic regression method for students' performance for prediction based on the data collected from the available students' information, financial information and learning management systems. Wolff et al. in [8], applied a decision-tree technique to detect the students with at risk academic situation. Jai Ruby & David in [16] used different data mining techniques for students' performance prediction. They found that the neural network technique outperforms the other data mining techniques and achieved heights student performance prediction of 74.8%

In [17], Abu Naser et al., used the Artificial Neural Network (ANN) model for the prediction of the performance of Engineering faculty students, they considered the students scores in the courses, number of passed credits and the cumulative grade point average as factors to evaluate the student performance, the ANN model correctly predict the student performance with accuracy of 80%

Ajith et al. in [18], proposed a rule mining framework for the evaluation of the performance of the students based on the Association Rules. Association Rules utilized to explore the student dataset for obtaining significant information for the student performance evaluation.

3. The Proposed Approach For Student Performance Prediction

The proposed approach mainly consists of two parts. In the first part, the results of the students in the previous exams are preprocessed by normalizing the student results to enhance the accuracy and the efficiency of the prediction. In the second part, the proposed approach is based on ANFIS is applied to predict the students’ expected performances in the next semester. Fig. 1 shows the proposed approach for student performance prediction.

3.1 Data preprocessing

The used dataset consists of 100 samples for computer science students, there are 5 compulsory subjects in the dataset, scores gained by students in these compulsory subjects are in the range from 0-100. The first step of the proposed approach is the preprocessing of the dataset. This step is potentially important to decrease error during the learning process, and to obtain more clear inputs for data mining approaches [19].

The normalization is accomplished by computing the mean and standard deviation of the data samples. Equation 1 provides the normalized sample \( X_i \) using the sample mean \( \mu \), and the standard deviation \( \sigma \) of the data sample.

\[
X_i = \frac{X_i - \mu}{\sigma}
\] (1)

3.2 The Adaptive Neuro Fuzzy Inference System

ANFIS gathers the advantages of fuzzy logic and neural network [20]. It has been successfully used in classification and prediction tasks. A simple architecture for ANFIS consists of two inputs (x and y), two fuzzy rules and single output (k), the Takagi and Surgeons' fuzzy rules used in ANFIS can be expressed as follows:

\[
R_1 = IF \ x \ is \ M_1 \ and \ y \ is \ N_1, \ then \ k_1 = c_1x + d_1y + z_1
\] (2)

\[
R_2 = IF \ x \ is \ M_2 \ and \ y \ is \ N_2, \ then \ k_2 = c_2x + d_2y + z_2
\] (3)

\( x \) and \( y \) denote the inputs to the ANFIS system, \( M_i \) and \( N_i \) represent the fuzzy sets, \( k_i \) represent the outputs of the fuzzy rule, while \( c_i, d_i \) and \( z_i \) represent the parameters
specified through the training process. Fig. 2 shows the ANFIS which consists of 5 layers, each layer performs certain function. A membership function is represented by a curve that shows exactly how the points in the input data are accurately be mapped to a membership degree in a range between the two values 0 and 1 [21]. In this paper, three varieties of membership functions that are utilized to define the fuzzy inference system parameters, viz. Gaussian MF (gaussmf), triangular MF (trimf) and Generalized Bell MF (gbellmf). The three membership functions are defined as follows:

\[
gbellmf(x, a, b, c) = \frac{1}{1 + \left| \frac{x-a}{b-a} \right|^c}
\]

\[
trimf(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0 \right)
\]

\[
gaussmf(x, c, \sigma) = \exp\left(-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2\right)
\]

Layer 1: The membership function for each fuzzy rule is computed in this layer.

Layer 2: Finds the firing strength for each rule which is similar to the membership value obtained in layer 1

Layer 3: Normalizes the firing strength for each fuzzy rule.

Layer 4: Computes the reasoning result of a rule.

Layer 5: aggregate the results of all fuzzy rules performed Layer 4.

Fig. 2 Different layer in the ANFIS.

In this paper, the dataset is partitioned into two subsets (80% as training data and 20% as testing data). We used ANFIS structure which consists of 5 inputs and one single output. The five inputs represent the student's scores in 5 courses and the GPA was the output. The fuzzy rules are constructed from the membership functions (MF). The ANFIS model consists of three membership functions to build the ANFIS model.

3.3 Performance Evaluation of the Proposed Model

The Root Mean Square Error is used to measure the accuracy of the students' performance prediction by finding the difference between performance of the actual observed data values and the ones predicted by the model. The RMSE is computed as follows:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{obs,i} - x_{model,i})^2}{n}}
\]

where \(x_{obs}\) denotes the observed performance and \(x_{model}\) is the predicted performance.

4. Experimental Results and Discussion

Various experiments were conducted using the three ANFIS models (ANFIS-GaussMF, ANFIS-TriMF and ANFIS-GbellMF). The ANFIS models were trained using the training data subset, while the test data subset was used to evaluate the prediction accuracy of the trained ANFIS models. In addressing the prediction of students' performance, student scores in the previous semester were used as inputs to the ANFIS models. The optimal ANFIS model will be nominated based on the testing RMSE values for different epoch numbers. The performance of the three ANFIS models based on three types of membership functions (GaussMF, TriMF, and GbellMF) is evaluated by varying the number of training epochs from 100 to 1000 as shown in table 1. The table depicts results which reflect that ANFIS model showed significant variation in its performance based on the type of membership function in use. The testing RMSE values for the three ANFIS models ranged from 0.193 to 0.777. The ANFIS-GbellMF model achieved the best students' performance prediction with RMSE of 0.193 in 1000 epochs, followed by the ANFIS-GaussMF model with RMSE of 0.216 in 1000 epochs respectively. The ANFIS-TriMF performs poorly and predicts the students' performance with RMSE of 0.754 which is less accurate than the other models.
Table 1: Training and testing RMSE using different types of membership function with different epochs

<table>
<thead>
<tr>
<th>No of Epochs</th>
<th>Membership function Type</th>
<th>Root mean square error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training Error</td>
</tr>
<tr>
<td>100</td>
<td>Gauss</td>
<td>0.624</td>
</tr>
<tr>
<td>100</td>
<td>TRIMF</td>
<td>0.027</td>
</tr>
<tr>
<td>100</td>
<td>Gbell</td>
<td>0.708</td>
</tr>
<tr>
<td>200</td>
<td>Gauss</td>
<td>0.172</td>
</tr>
<tr>
<td>200</td>
<td>TRIMF</td>
<td>0.026</td>
</tr>
<tr>
<td>200</td>
<td>Gbell</td>
<td>0.166</td>
</tr>
<tr>
<td>300</td>
<td>Gauss</td>
<td>0.152</td>
</tr>
<tr>
<td>300</td>
<td>TRIMF</td>
<td>0.025</td>
</tr>
<tr>
<td>300</td>
<td>Gbell</td>
<td>0.149</td>
</tr>
<tr>
<td>400</td>
<td>Gauss</td>
<td>0.15</td>
</tr>
<tr>
<td>400</td>
<td>TRIMF</td>
<td>0.02</td>
</tr>
<tr>
<td>400</td>
<td>Gbell</td>
<td>0.147</td>
</tr>
<tr>
<td>500</td>
<td>Gauss</td>
<td>0.149</td>
</tr>
<tr>
<td>500</td>
<td>TRIMF</td>
<td>0.02</td>
</tr>
<tr>
<td>500</td>
<td>Gbell</td>
<td>0.147</td>
</tr>
<tr>
<td>600</td>
<td>Gauss</td>
<td>0.152</td>
</tr>
<tr>
<td>600</td>
<td>TRIMF</td>
<td>0.02</td>
</tr>
<tr>
<td>600</td>
<td>Gbell</td>
<td>0.147</td>
</tr>
<tr>
<td>700</td>
<td>Gauss</td>
<td>0.148</td>
</tr>
<tr>
<td>700</td>
<td>TRIMF</td>
<td>0.02</td>
</tr>
<tr>
<td>700</td>
<td>Gbell</td>
<td>0.147</td>
</tr>
<tr>
<td>800</td>
<td>Gauss</td>
<td>0.147</td>
</tr>
<tr>
<td>800</td>
<td>TRIMF</td>
<td>0.019</td>
</tr>
<tr>
<td>800</td>
<td>Gbell</td>
<td>0.146</td>
</tr>
<tr>
<td>900</td>
<td>Gauss</td>
<td>0.147</td>
</tr>
<tr>
<td>900</td>
<td>TRIMF</td>
<td>0.019</td>
</tr>
<tr>
<td>900</td>
<td>Gbell</td>
<td>0.146</td>
</tr>
<tr>
<td>1000</td>
<td>Gauss</td>
<td>0.147</td>
</tr>
<tr>
<td>1000</td>
<td>TRIMF</td>
<td>0.018</td>
</tr>
<tr>
<td>1000</td>
<td>Gbell</td>
<td>0.146</td>
</tr>
</tbody>
</table>

Figure 3 shows the training RMSE for the three ANFIS models. It is noticeable that for the three models when the number of epochs reaches about 200 iterations, the training RMSE almost reaches a stable value and that means more training to the model will not yield significant results in the prediction accuracy of the ANFIS model and may cause over-fitting of the ANFIS model.

Fig. 3 Training RMSE for the three models with different numbers of training epochs.

Fig.4 clearly shows that the testing RMSE values of the ANFIS-GbellMF is lower than that of ANFIS-GaussMF, while the ANFIS-TriMF achieved the worst testing RMSE.

Figure 4: Testing RMSE for the three models with different numbers of training epochs.

5. Conclusions

Student’s performance prediction is important to assist students in improving their academic performance. This paper presented an adaptive neuro fuzzy inference system for students' performance prediction. The used ANFIS models for student performance prediction utilized three types of membership functions and the backpropagation algorithm for training. ANFIS model based on the generalized bell-shaped membership function (ANFIS-GbellMF) provided superior prediction of the student's performance with a RMSE as low as 0.193. Future work will utilize different data mining approaches and will incorporate larger dataset to achieve results that are more accurate. Data mining approaches such as Clustering can
be applied to the dataset before using the ANFIS to improve the prediction capabilities of the models as well as their accuracy.

**Acknowledgment**

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under Grant No. (160-830-1437). The authors, therefore, gratefully acknowledge the DSR technical and financial support.

**References**


