A Comparative Study of Predictions of Students’ Performances in Offline and Online Modes using Adaptive Neuro-Fuzzy Inference System

Sudhindra B. Deshpande, Kiran K.Tangod, Neelkanth V. Karekar, Pandurang S. Upparmani  
Department of Information Science and Engineering, KLS, Gogte Institute of Technology  
Belagavi-06, Karnataka, India

Abstract—Predicting a student’s performance can help educational institutions to support the students in improving their academic performance and providing high-quality education. Creating a model that accurately predicts a student’s performance is not only difficult but challenging. Before the pandemic situation students were more accustomed to offline i.e., physical mode of learning. As covid-19 took over the world the offline mode of education was totally disturbed. This situation resulted into the new beginning towards online mode of teaching over the Internet. In this article, these two modes are analysed and compared with reference to students’ academic performances. The article models a predicting academic performance of students before covid i.e., physical mode and during Covid i.e., online mode, to help the students to improve their performances. The proposed model works in two steps. First, two sets of students’ previous semester end results (SEE) i.e., after offline mode and after online mode, are collected and pre-processed using normalizing the performances in order to improving the efficiency and accuracy. Secondly, Adaptive Neuro-Fuzzy Inference System (ANFIS) is applied to predict the academic result performances in both learning modes. Three membership functions gaussian (Gausmf), triangular (Trimf) and gaussian-bell (Gbellmf) of ANFIS are used to generate the fuzzy rules for the prediction process proposed in this paper.

Keywords—ANFIS; fuzzy systems; online learning; e-learning; classroom learning; fuzzy rules; predictions; adaptive neuro-fuzzy inference system; education technology; distance education

I. INTRODUCTION

The Internet has made "Education" as the most exciting domains available in today’s society, especially with the interconnectivity of smart and portable devices. Distance and e-Learning has grown in popularity as computer technology has advanced, networks have improved, and smartphones have become more accessible. Technology has transformed general domains available in today’s society, especially with the interconnectivity of smart and portable devices. Distance and e-Learning has grown in popularity as computer technology has advanced, networks have improved, and smartphones have become more accessible. Technology has transformed general learning process takes place in a classroom, where a teacher uses lectures, whiteboards, slide projectors, and group discussions to communicate knowledge. To improve their learning skills, students take notes, write tests and assignments, and express doubts and questions. Students must also take tests and assessments on paper and teachers assess performance and knowledge of students in terms of marks. A variety of technological media, smart mobile devices and wireless communication have revolutionized the education system [1]. The e-Learning may happen through a variety of processes and applications such as computer-based training, digital collaborations, web-based training, and virtual classrooms [2].

Whether it is classroom learning or e-learning, one very important aspect is performance evaluation of students. Educational institutes are attempting to include student performance prediction into their educational procedures in order to help the students to improve themselves. Traditional learning modes can be enhanced to online modes by increasing portability and accessibility of mobile networks. Mobile devices can make it possible to capture the user-created content and also can serve as a powerful data collection tool [3]. Gayeski, D., & Russell, J. D. [4] have proposed that evaluations can be done in several ways. The teacher needs to make evaluations to advise students to make decisions and to write recommendations to potential employers. Data mining and machine learning techniques have been in use in HEIs to guide and improve their students’ performance by using and exploring the data available in the education domain [5] [6] [7].

Artificial neural networks, Bayesian classifiers, and support vector machine algorithms are some of the techniques used to classify data [8]. Various techniques including classification, clustering, visualization, and regression have been utilized to extract hidden information from educational databases [9]. Classification is a technique for assigning data values to predetermined classes with the goal of predicting the target class for each data value [10][11].

Education data mining peruses extensive educational datasets for important data that can be analyzed deeper. Many educational activities, such as learners' performance prediction, will be supported by the smeared data, allowing teachers to identify potential knowledge about students [12][13]. The Neuro-fuzzy inference system (NFIS) has been effectively implemented in a variety of applications, including control and classification [14]. NFIS is a machine learning tool that integrates fuzzy logic reasoning with the learning capabilities of neural networks, thereby addressing the drawbacks of both neural networks and fuzzy systems when used separately [15]. A tutoring system based on Fuzzy rules [16] which implements a Fuzzy Logic inference engine that can manage different learning activities of students. A tutoring system that uses Fuzzy Logic is demonstrated to monitor the cognitive capability of each student [17].
In this article, ANFIS is applied to evaluate students’ academic performance in order to assist them in improving their grades both in online and offline modes. The suggested ANFIS-based solution intelligently mixes fuzzy logic's capability reasoning with neural networks' learning abilities. The use of ANFIS for reliable student performance prediction in online and offline mode, is one of the paper's primary contributions.

II. LITERATURE WORK

A rule mining approach is introduced for evaluating student performance based on the Association Rules [18]. To collect crucial information for the student performance evaluation, Association Rules were used to examine the student dataset. Various data mining techniques are employed to predict student performance. They discovered that the neural network technique outperformed the other data mining strategies, with a prediction accuracy of 74.8% for student performance [19]. Abu Naser et al. used the Artificial Neural Network (ANN) model to predict the performance of Engineering faculty students [20]. They used the students' course scores, number of passed credits, and cumulative grade point average as factors to evaluate the student performance, and the ANN model correctly predicted the performance of students with an accuracy of 80%. A Neuro-Fuzzy model was proposed to analyze errors of high school learners’ by collecting data using course-related simulation tools namely vectors in mathematics and physics. The system was tested using simulated learner data with different categories of knowledge level. These learners' behaviors correspond to fuzzy values [21]. A model of feed-forward Neural Networks was also trained for error classification purposes. Barber and Sharkey employed the logistic regression method for predicting student performance based on data collected from available students’ information, financial data, and learning management systems [22].

A Fuzzy learner model was discussed for student evaluation during learning activities. The procedure is to imitate a human teacher in the classroom. Fuzzy Logic is proposed to track interaction between the tutor and the students and to handle inaccurate information using the ability of Fuzzy Logic [23]. This leads to more accurate answers by students to improve the learning environment.

Xenos, M. [24] has proposed a model of the Bayesian Network to support educational tutors in making decisions under uncertain conditions. The system is implemented with 800 learners studying an informatics course. The system is designed to model learner behavior to predict the success of decision-making by tutors. Fadi R.S [25] has discussed the early stages of data mining in academics and highlighted the potential of data mining in e-Learning and suggested various data mining tools that can be beneficial for e-Learning.

III. MOTIVATION

As the Covid-19 pandemic has affected the activities of higher education institutions (HEI) to promote the protection of teachers, staff, and students during the public health emergency. Institutions had to cancel all face-to-face lectures, including labs and other learning experiences, and determined that teachers would completely convert their courses to emergency online learning, and reduced contacts to prevent the spread of Virus. So, to strengthen the learning community in this pandemic situation online learning has been promoted. But for the first-time online learning mode is actively used in the education. A comparative analysis is required between online and physical modes of teaching to find out how the student community is adjusting to the online mode of teaching and whether it is successful and helpful even after the Covid pandemic.

IV. THE PROPOSED METHODOLOGY

The proposed methodology is split into four steps. First step deals with data collection and processing; in second part, students’ results in the classroom mode are analyzed using ANFIS; in third part, students’ results in the online mode are analyzed using ANFIS; lastly, both modes are compared for their performances to evaluate which of the teaching mode is effective in producing good results. Fig. 1 depicts the proposed methodology applied. An Adaptive Neuro-Fuzzy Inference System [26][27] is a sort of artificial neural system that depends on the Takagi–Sugeno fuzzy inference system. It incorporates both neural systems and fuzzy rationale standards; it can catch the advantages of both in a solitary structure. Its inference system compares to an arrangement of fuzzy IF-THEN rules that have learning ability to surmised nonlinear functions.

A. Data Collection and Processing

The dataset used consists of 210 samples for Information science students consisting of students’ scores in four core subjects and two laboratories ranging from 0 to 100. There are two sets of datasets which are collected. One set belongs to the 1st, 3rd and 5th semester students’ semester end examination (SEE) results obtained through classroom learning mode and before the pandemic. Second set consists of SEE results of same students but studied in 2nd, 4th and 6th semester during the pandemic through online mode learning. The preparation of the dataset is the initial stage in the proposed approach. This stage could be crucial for reducing error throughout the learning process and obtaining more precise inputs. Equation (1) is used to normalize datasets using sample mean μ and standard deviation σ.

$$X_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

B. ANFIS Architecture

“Takagi–Sugeno fuzzy inference framework” is used in ANFIS architecture which is an adaptive system that utilizes a directed learning algorithm. Fig. 2 demonstrates the architecture of ANFIS that consists of two sources of information “x” and “y”, and one yield or output “f”.

Takagi–Sugeno model follows the two variants of IF-THEN, are

RULE 1 = If ‘x’ is A1 and ‘y’ is B1 Then f1 = p1x+q1x+r

RULE 2 = If ‘x’ is A2 and ‘y’ is B2 Then f2 = p2y+q2y+r2

where A1, A2, and B1, B2 are the membership elements of each info ‘x’ and ‘y’, while p1, q1, r1 and p2, q2, r2 are linear parameters to a limited extent Then (subsequent part) of Takagi–Sugeno fuzzy inference model.
Fig. 1. Proposed Methodology of the Work Carried Out.

Fig. 2. ANFIS Architecture.

Fig. 3 shows the functions of five layers of ANFIS. The first layer and fourth layer consist of adaptive nodes, while the remaining layers consist of fixed nodes. ANFIS tool in MATLAB provides a Grid partitioning algorithm. The algorithm generates a single-output Sugeno-type FIS by using grid partitioning on the data. genfis1 generates a Sugeno-type FIS structure used as initial conditions (initialization of the membership function parameters) for ANFIS training. It generates input membership functions by uniformly partitioning the input variable ranges, and creates a single-output Sugeno fuzzy system. The fuzzy rule base contains one rule for each input membership function combination.

The membership function is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. They characterize fuzziness as whether the elements in fuzzy sets are discrete or continuous. The membership functions "trimf, gaussmf and gbellmf" - are explained:

"Triangular function: defined by a lower limit a, an upper limit b, and a value m, where a < m < b, as given by equation (2)."
\[ \mu_A(X) = \begin{cases} 0 & x = a \\ \frac{m-a}{b-a} & b - x \\ \frac{b-m}{c-m} & c > x \end{cases} \]

Gaussian function: defined by a central value \( m \) and a standard deviation \( k > 0 \) as given by the equation (3).

\[ \mu_A(X) = e^{-\frac{(x-m)^2}{2k^2}} \]

(3)

“The generalized bell function: gbellmf depends on three parameters \( a, b, \) and \( c \) given by the equation (IV) where the parameter \( b \) is usually positive and the parameter \( c \) locates the center of the curve.”

\[ f(x; a, b, c) = \frac{1}{1 + \left| \frac{x-a}{b} \right|^{2b}} \]

(4)

The datasets are partitioned into 70:30 i.e., 70% training data and 30% testing data. For both models of ANFIS, scores of six subjects are considered as input variables and SGPA as output variable. Three membership functions, gaussian (Gaussmf), triangular (Trimf) and gaussian-bell (Gbellmf) of ANFIS are used to generate the fuzzy rules for both the models.

C. Performance Evaluation

The Root Mean Square Error is a metric for determining the accuracy of a student’s performance prediction by comparing the actual observed data values to the ones predicted by the model. The formula in equation (5) is used to calculate the RMSE:

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

(5)

V. RESULTS AND DISCUSSIONS

Two ANFIS models using three mfs are built for the experimentation; one for classroom/offline mode of learning and second is for online mode of learning. The training data subset is used to train the ANFIS models, while the test data subset is employed to assess the trained ANFIS models’ prediction accuracy. Student scores from the previous semester are taken as inputs to the ANFIS models in order to forecast student performances. The testing RMSE values for different epoch numbers will be used to designate the best ANFIS model in both learning modes. By varying the number of training epoch from 50 to 900, the performance of the three ANFIS models based on three types of membership functions (GaussMF, TriMF, and GbellIMF) is examined.

A. ANFIS for Classroom / Offline Mode

Table I displays the training and testing RMSEs for offline mode learning approach for the three types of mfs.

Table I. TRAINING AND TESTING RMSEs OF GAUSSMF, GBELLMF AND TRIMF WITH DIFFERENT EPOCHS FOR OFFLINE MODE

<table>
<thead>
<tr>
<th>MFs</th>
<th>GaussMF</th>
<th>GbellMF</th>
<th>TriMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Epochs</td>
<td>Train g error</td>
<td>Test g error</td>
<td>Train g error</td>
</tr>
<tr>
<td>50</td>
<td>0.73</td>
<td>0.792</td>
<td>0.78</td>
</tr>
<tr>
<td>100</td>
<td>0.659</td>
<td>0.669</td>
<td>0.756</td>
</tr>
<tr>
<td>150</td>
<td>0.61</td>
<td>0.77</td>
<td>0.698</td>
</tr>
<tr>
<td>200</td>
<td>0.598</td>
<td>0.7</td>
<td>0.645</td>
</tr>
<tr>
<td>250</td>
<td>0.245</td>
<td>0.566</td>
<td>0.388</td>
</tr>
<tr>
<td>300</td>
<td>0.179</td>
<td>0.456</td>
<td>0.356</td>
</tr>
<tr>
<td>350</td>
<td>0.171</td>
<td>0.423</td>
<td>0.355</td>
</tr>
<tr>
<td>400</td>
<td>0.169</td>
<td>0.389</td>
<td>0.289</td>
</tr>
<tr>
<td>450</td>
<td>0.168</td>
<td>0.389</td>
<td>0.219</td>
</tr>
<tr>
<td>500</td>
<td>0.169</td>
<td>0.378</td>
<td>0.17</td>
</tr>
<tr>
<td>600</td>
<td>0.168</td>
<td>0.378</td>
<td>0.169</td>
</tr>
<tr>
<td>700</td>
<td>0.168</td>
<td>0.345</td>
<td>0.17</td>
</tr>
<tr>
<td>800</td>
<td>0.168</td>
<td>0.336</td>
<td>0.17</td>
</tr>
<tr>
<td>900</td>
<td>0.168</td>
<td>0.334</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Fig. 5(a) depicts the RMSE values of training for all the three-membership function. It can be observed that at 400 epoch iterations, training RMSE of gaussmf almost reaches a stable value. At 500 epoch iterations, training RMSE of gbellmf and trimf reach stable values. Further increase in the number of epochs may not yield any significant results in predictions causing overfitting of the models. Fig. 5(b) depicts that testing RMSE values of gaussmf is less than the gbellmf and trimf, meaning ANFIS-gaussmf model with least testing RMSE value of 0.334 predicts accurately. And trimf with 0.877 testing RMSE performs the worst in predictions.

A. ANFIS for Online/e-Learning Mode

Table II displays the training and testing RMSEs for online mode learning approach for the three types of mfs of ANFIS.

Fig. 6 depicts MSE values of training and testing against the number of epochs for all the three membership functions for online mode of learning; and trimf behaviour is found to be worst as the testing RMSE is very high with 0.878 against a very low training error of 0.025. gaussmf performs good in comparison to gbellmf and trimf with 0.334 testing RMSE.

Fig. 7(a) depicts the RMSE values of training for all the three-membership function. It can be observed that at 500 epoch iterations, training RMSE of all the three mfs reach stable values. Further increase in the number of epochs may not yield any significant results in predictions causing overfitting of the models. Fig. 7(b) depicts that testing RMSE values of gaussmf and gbellmf are almost same and are very low compared trimf, meaning ANFIS-gaussmf model and ANFIS-gbellmf model predict accurately. And trimf with 0.877 testing RMSE performs the worst in predictions.
Fig. 4. Training and Testing RMSEs vs the Number of Epochs (a) Guassmf (b) Gbellmf (c) Trimf, for Offline Mode Predictions.

Fig. 5. Training RMSEs of 3 Mfs and Testing RMSEs of 3 Mfs for Offline Mode Predictions.
Fig. 6. Training and Testing RMSEs vs the Number of Epochs (a) Gaussmf (b) Gbellmf (c) Trimf, for Online Mode Predictions.

Fig. 7. Training RMSEs of 3 Mfs and Testing RMSEs of 3 Mfs for Online Mode Predictions.
Training and testing RMSE of ANFIS-gaussmf, gbellmf and trimf with different epochs for online mode

<table>
<thead>
<tr>
<th>No. of Epochs</th>
<th>Training Error</th>
<th>Testing Error</th>
<th>Training Error</th>
<th>Testing Error</th>
<th>Training Error</th>
<th>Testing Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.789</td>
<td>0.779</td>
<td>0.801</td>
<td>0.773</td>
<td>0.0697</td>
<td>0.989</td>
</tr>
<tr>
<td>100</td>
<td>0.735</td>
<td>0.725</td>
<td>0.755</td>
<td>0.756</td>
<td>0.056</td>
<td>0.892</td>
</tr>
<tr>
<td>150</td>
<td>0.689</td>
<td>0.667</td>
<td>0.667</td>
<td>0.678</td>
<td>0.059</td>
<td>0.899</td>
</tr>
<tr>
<td>200</td>
<td>0.635</td>
<td>0.655</td>
<td>0.678</td>
<td>0.689</td>
<td>0.058</td>
<td>0.891</td>
</tr>
<tr>
<td>250</td>
<td>0.489</td>
<td>0.58</td>
<td>0.511</td>
<td>0.599</td>
<td>0.056</td>
<td>0.881</td>
</tr>
<tr>
<td>300</td>
<td>0.389</td>
<td>0.478</td>
<td>0.39</td>
<td>0.467</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>350</td>
<td>0.287</td>
<td>0.334</td>
<td>0.392</td>
<td>0.467</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>400</td>
<td>0.256</td>
<td>0.311</td>
<td>0.356</td>
<td>0.411</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>450</td>
<td>0.221</td>
<td>0.297</td>
<td>0.289</td>
<td>0.356</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>500</td>
<td>0.221</td>
<td>0.297</td>
<td>0.284</td>
<td>0.297</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>600</td>
<td>0.221</td>
<td>0.297</td>
<td>0.284</td>
<td>0.297</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>700</td>
<td>0.221</td>
<td>0.284</td>
<td>0.289</td>
<td>0.291</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>800</td>
<td>0.221</td>
<td>0.284</td>
<td>0.289</td>
<td>0.291</td>
<td>0.057</td>
<td>0.878</td>
</tr>
<tr>
<td>900</td>
<td>0.22</td>
<td>0.284</td>
<td>0.289</td>
<td>0.291</td>
<td>0.057</td>
<td>0.878</td>
</tr>
</tbody>
</table>

B. Comparison of Offline and Online ANFIS Models

Table III depicts training and testing RMSE of all three ANFIS mfs for epoch 900 for offline and online learning modes referring to the Tables.

<table>
<thead>
<tr>
<th>Learning Modes</th>
<th>MFs</th>
<th>Gaussian MF</th>
<th>Gbell MF</th>
<th>Tri MF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Epochs</td>
<td>Training Error</td>
<td>Testing Error</td>
<td>Training Error</td>
</tr>
<tr>
<td>offline mode</td>
<td>900</td>
<td>0.168</td>
<td>0.334</td>
<td>0.171</td>
</tr>
<tr>
<td>online mode</td>
<td>900</td>
<td>0.22</td>
<td>0.284</td>
<td>0.289</td>
</tr>
</tbody>
</table>

Training and testing RMSE for ANFIS-gaussmf are 0.22 and 0.284; and training and testing RMSE for ANFIS-gbellmf are 0.289 and 0.291. These gaussmf and gbellmf RMSE values of online learning mode show that training and testing RMSEs are almost same without a large variation. And offline mode’s ANFIS-gaussmf testing RMSE (0.334) is larger than the online mode’s testing RMSE ANFIS-gaussmf (0.284); means predicting the performances of student results in online mode is superior and accurate compared to offline predictions.

VI. CONCLUSION

Before the Covid-19 students used to learn engineering subjects through physical i.e., classroom or offline mode by attending classes. At the end of semester students are assessed to check their performances by writing semester end examinations (SEE). As Covid-19 pandemic begun, these offline classes were totally suspended and e-Learning or online mode of learning using Internet and mobile devices took over the education in a different direction. Online mode in the education domain has initiated a new paradigm shift. So, the students learned through online mode during the pandemic and also SEE were conducted after the semester end. So, there is a need to analyse and predict which mode of education (online or offline) impacted the student’s learning process and improve the learning curve by increasing the performances. This article illustrated the ANFIS approach to model and predict the students’ performances obtained through both modes of learning. Firstly, ANFIS for offline mode was built and found that ANFIS-gaussmf predictions are better. Secondly, ANFIS for online mode was built and found that ANFIS-gaussmf and gbellmf, both predicted the same and well performed. In both, ANFIS-trimf performed very worst. Finally, ANFIS-online mode performed very well compared to ANFIS-offline mode i.e., predicting the performances of student results in online mode is superior and accurate compared to offline predictions.

REFERENCES


