International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Volume 1 Issue 6 (July 2014) http://ijirae.com

The role of Dataset in training ANFIS System for Course Advisor

V.Vaidhehi
Department of Computer Science, Christ University

Abstract— ***Adaptive Network based Fuzzy Inference System (ANFIS) is used in the field of decision making to help the students to choose the best course according to his/her requirements. The structure of ANFIS system and the datasets used to train the system play a vital role in evaluating the performance of the system. This paper is based on the design of Sugeno type ANFIS with grid partitioning and the usage of different datasets to train the system using MATLAB. Results demonstrate that proper dataset is needed for training the ANFIS model***

Keywords— course advisor, ANFIS, dataset, grid partitioning, neuro fuzzy system, Sugeno, MATLAB

I. INTRODUCTION

Technology based course advisory systems help the student to choose the best course through knowledge base and reasoning ability. Decision support systems can be designed with artificial neural network [7]. Design of neuro fuzzy system [8] shows its application in the real world problems. Neuro Fuzzy Model for the student domain with feed forward architecture with five layers of neurons and four connections is discussed in [12]. MATLAB implementation of neuro fuzzy system is discussed in [13]. Neural network construction for precollege students is discussed in [14]. Decision making using fuzzy based system is studied in [15].

Adaptive Network based Fuzzy Inference System ANFIS is implemented as a Sugeno fuzzy inference system. ANFIS system allows the user to choose or modify the parameters of the membership functions based on the data. The parameters are adjusted automatically by the neuro adaptive learning techniques like back propagation algorithm or hybrid method (which is a combination of back propagation and least squares method). These techniques allow the fuzzy inference system to learn information about the data set. During the learning process, the parameters of the membership functions will be changed. The computations of these parameters can be controlled by using the optimization procedure which is defined by the sum of squared difference between actual and desired outputs. Sugeno systems are more compact and computationally efficient representation than a Mamdani system.

II. STRUCTURE OF ANFIS SYSTEM

Structure of the Sugeno model is designed in such a way that the input is mapped to input membership function, the input membership function is mapped to rule, then the rule is mapped to output membership function and then the output membership function is mapped to the output. Thus the system takes five layers.

Each node in the first layer generates a membership grade. Each node in the second layer calculates the firing strength of the rule. Each node in the third layer calculates the ratio of the ith rule's firing strength to the total of all firing strength. Each node in the fourth layer is an adaptive node which maps to the output membership functions. The node in the fifth layer gives the overall output. This structure is shown in figure 1.

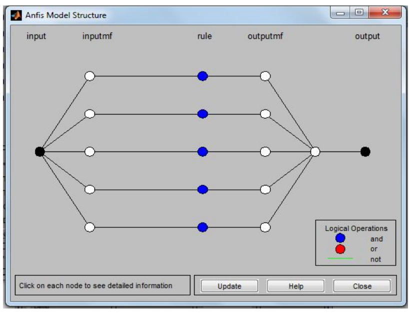

Figure 1 Structure of ANFIS.

III. ANFIS MODEL

To model ANFIS for any real world problems, Let n be the various possible inputs of the application. Let r be the effective inputs of the application. Then ncr =x has to be computed. X models has to be trained with a single pass of least square method. The one with minimum training error is selected for further training using Hybrid method. One epoch training of X model involves less computation than training a single model with X epochs. Identifying the structure of FIS, parameter selection for training are the factors affecting the system performance.

Model validation is the process by which the input data sets on which the FIS was not trained, are presented to the trained FIS model, to see the performance. When checking data is given, the FIS model is expected to have parameters associated with the minimum checking data model error. The model validation for models constructed using adaptive techniques is essential. The data set used for model validation should have the features of training data set and also significantly distinct from the training data set. A large amount of data is collected with all the necessary representative features, so that the process of selecting a data set for checking or testing purposes is made easier.

GRID Partitioning

Grid partition divides the data space into rectangular sub spaces called "grids" based on the number of membership functions and their types. The usage of grid partition in any application, is restricted because of the curse of dimensions. The number of fuzzy rules increases exponentially, when the number of input variables increases. For example, if there are m membership functions for every input variable and a total of n input variables for the problem, then the total number of fuzzy rules needed is m^n. Therefore, the grid partition is only suitable for applications with small number of input variables (generally less than 6). In this student course selection domain, the tenth average and four core subject marks in class XII are considered as input variables. Grid partitioning models can run accurately with a few number of membership functions. This requires less simulation time. Grid partitioning method achieves low error values by using a few number of membership functions. On the other hand, Subtractive clustering function produces accurate output values by using a large number of membership functions. This requires more simulation time. Thus it is reasonable to apply the grid partition to generate FIS structure.

Training

It is a learning process of the developed model. The model is trained till the results are obtained with minimum error. To design an ANFIS system for real world problems, it is essential to select the parameters for the training process. It is essential to have proper training and testing data sets. If the datasets are not selected properly, then the testing data set will not validate the model. If the testing data set is completely different from the training dataset, then the model cannot capture any of the features of the testing data. Then, the minimum testing error can be achieved in the first epoch. For the proper data set, the testing error decreases with the training proceeding until a jump point. Over fitting occurs when the training passes that point. The optimization methods are used to learn about the training data. During the learning process, the parameters of the memberships are updated. In MATLAB, the two ANFIS parameter optimization methods are hybrid (the default, mixed least squares and back propagation) and backpropa (back propagation). Error tolerance is used as training stopping criterion, which is related to the error size. The training will stop after the training data error remains within this tolerance.

Hybrid learning

The hybrid optimization method is a combination of least-squares and back propagation gradient descent method. There are two steps in Hybrid learning algorithm. They are

1. Forward pass
2. Backward pass

In the forward pass, premise parameters are fixed and least square estimation is used to update the consequent parameters. In the backward pass, consequent parameters are fixed and back propagation gradient descent method is used to update the premise parameters. By repeating the forward and backward passes, the premise and consequent parameters are identified for the FIS system. For this ANFIS model, the number of training epochs is 40 and training error tolerance is set to zero. The training process stops whenever the maximum epoch number is reached or the training error goal is achieved. In the training part, hybrid optimization method is faster and has closest results than the back propagation gradient descent optimization method.

IV. RESULTS

Training the ANFIS system with the training data set is shown in the figure1. The training error is the difference between the training data output value, and the output of the fuzzy inference system corresponding to the same training data input value, (the one associated with that training data output value). The training error records the root mean squared error (RMSE) of the training data set at each epoch. The ANFIS Editor GUI plots the training error versus epochs curve as the system is trained. Testing the trained FIS is shown in figure 2.The average testing error for the training data set is 0.034068.

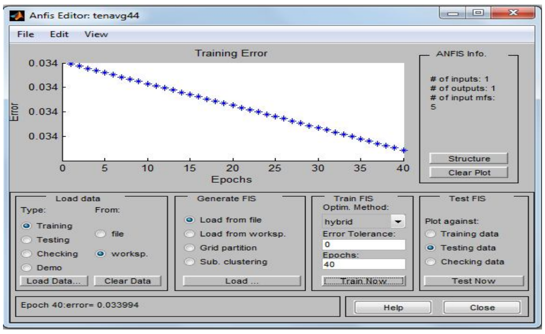

FIG 1: Training error

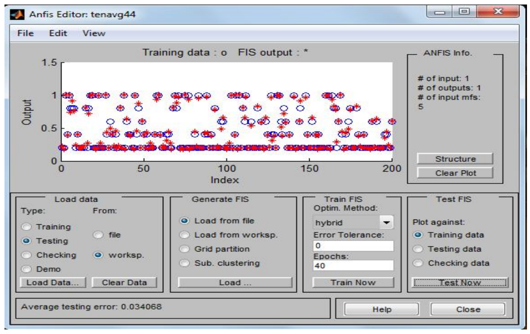

FIG 2: Testing the FIS with Training data set

Checking data is used for testing the generalization capability of the fuzzy inference system at each epoch. The checking data has the same format as that of the training data. This data set is used to validate the fuzzy inference model. This validation is done by applying the checking data to the model and then seeing how well the model responds to this data. When the checking data option is used using the ANFIS Editor GUI, the checking data is applied to the model at each training epoch. The FIS membership function parameters are computed using the ANFIS Editor GUI when both training and checking data are loaded. The checking data is similar enough to the training data that the checking data error decreases as the training begins. The checking error is the difference between the checking data output value, and the output of the fuzzy inference system corresponding to the same checking data input value, which is the one associated with that checking data output value. The checking error records the RMSE for the checking data at each epoch. The ANFIS Editor GUI plots the checking error versus epochs curve as the system is trained. The figure 3 shows the plot of training data and the checking data. The average testing error for the checking data set is 0.03382. When checking data is tested against the trained FIS, it looks satisfactory and it shown in figure 4.

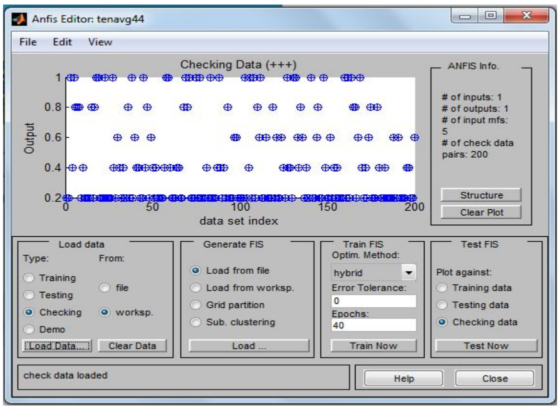

FIG 3: Training data set vs checking data set

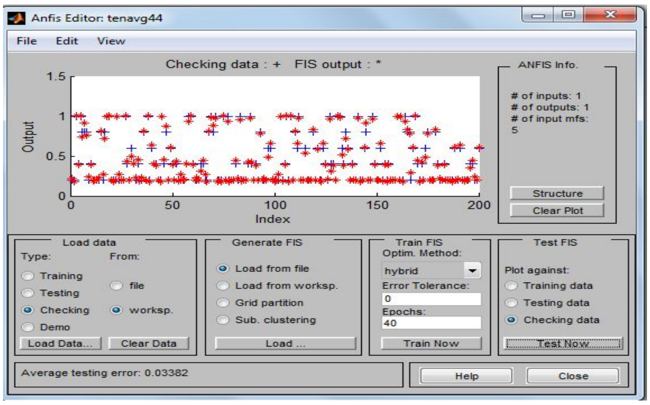

FIG 4: Testing the checking data set

The figure 5 shows the plot of training data and the test data which is slightly different from the one used to train the system. The figure 6 shows the testing of the data set which significantly different from training. The average testing error for the testing data set is 0.17268.

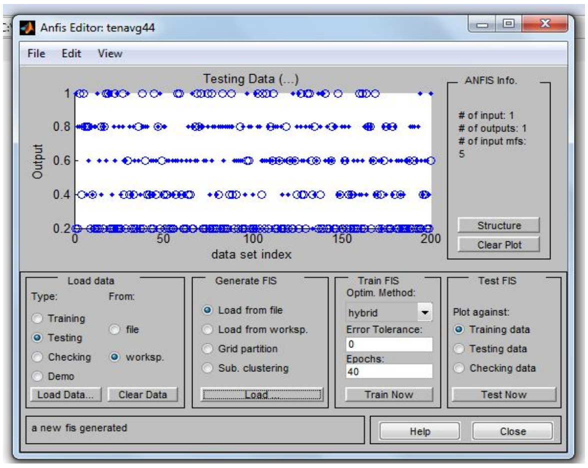

FIG 5: Testing data and training data set

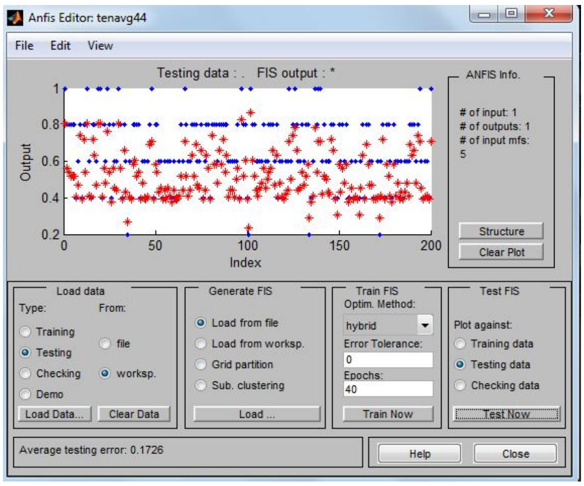

FIG 6: Testing of test data and training data set

When testing dataset is used to train the system, the average training error is 0.034031. It is clear that average training is reduced from the previous training and it is shown in figure 7. Testing the training data against the trained FIS is shown in figure 8.

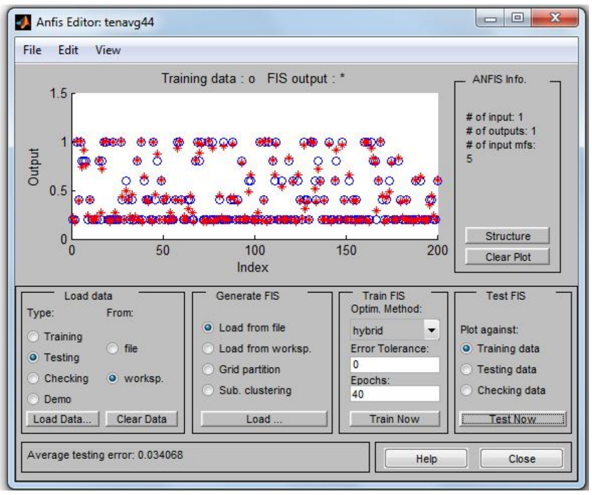

FIG 7 :Testing data as training data set

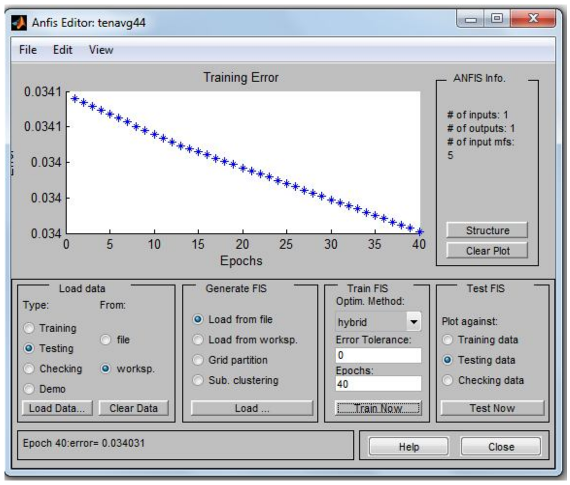

FIG 8 : Testing of training data set

V. Conclusion

This paper is based on design of Sugeno type ANFIS system with GRID paprtitoning technique. Learning of ANFIS is based on Hybrid learning. The implementation is done in MATLAB. Results indicate that the selection of datasets for training the ANFIS system is an important factor affecting the performance. Also performance of the ANFIS can be improved by using different training data sets. If the datasets used for testing is extremely different from the one of the training dataset, then the system fails to capture the essential features of the dataset. Therefore, the performance of the system is affected by the design of the datasets. The performance of the ANFIS can also be measured by varying the type of membership functions, partitioning techniques , learning methodology and type of ANFIS system.

Acknowledgment

This work is supported in part by the Christ University, Bangalore under Minor Research project.

References

[1] MA AI Ahmar, ”Prototype Student Advising Expert System Supported with an Object-Oriented Database” (IJACSA) International Journal of Advanced Computer Science and Applications, pp. 100-105, 2011.
[2] Wabwoba Masinde, “Students Selection for University Course Admission at the Joint Admissions Board (Kenya) Using Trained Neural Networks”, Journal of information technology education, volume 10, 2011.
[3]Feghali, T., Zbib, I., & Hallal, S, “A Web-based Decision Support Tool for Academic Advising.”, Educational Technology & Society, 14 (1), pp. 82–94,2011
[4] Fawzi Albalooshi , Safwan Shatnawi, ”HE-Advisor: A Multidisciplinary Web-Based Higher Education Advisory System”, Global Journal of Computer Science and Technology, 10(7), pp 37-40,2010.
[5] Emebo Onyeka, Daramola Olawande, Ayo Charles , “CAES: A Model of an RBR-CBR Course Advisory Expert System”, Information society,2010.
[6] Mahdi Hassani Goodarzi , Vahid Rafe, ”Education Advisor System implemented by Web based Fuzzy Expert Systems” , Asian Journal of Information Technology, 11: pp. 77-82, 2012.
[7] Priti Srinivas Sajja ,”Type-2 Fuzzy User Interface for Artificial Neural Network based Decision Support System for Course Selection”, IJCIR, vol 2 ,n2, 2008.
[8] R. ElHamdi, M Njah, M Chtourou, ”A Hybrid Evolutionary Design of Neuro-Fuzzy Systems”, 7th International Multi-Conference on Systems, Signals and Devices, 2010.
[9] Xin Wang, Fang Yuan, ”Course Recommendation by Improving BM25 to Identity Students’ Different Levels of Interests in Courses“, International Conference on New Trends in Information and Service Science, 2009.
[10] Kamal Taha, “Automatic Academic Advisor”, 8th International Conference on Collaborative Computing: Networking, Applications and Work sharing , October 14-17, 2012.

[11] Hoi-Fei Kwok, Derek A. Linkens, Mahdi Mahfouf, and Gary H. Mills, "SIVA: A Hybrid Knowledge-and-Model-Based Advisory System for Intensive Care Ventilators" IEEE Transactions on Information Technology and medicine, VOL. 8, NO. 2, June 2004.

[12] Khalid Isa, Shamsul Mohamad, Zarina Tukiran, "Development of INPLANS: An Analysis on students' performance using Neuro Fuzzy", AIA,2007.

[13] Urvashi Rahul Saxena, S.P. Singh, "Integration of Neuro Fuzzy systems to develop Intelligent Planning systems for predicting students' performance",

[14] Jesse L Bernley," Neural networks for precollege students", IJCNN '01, Page(s): 538 - 539 vol.1

[15] M.A. Owais ,"Subjective decision making using Type 2 fuzzy logic advisor", Information and Communication Technologies, 2009, Page(s): 127 – 133

[16] Oge marques, Xundong Ding, Sam Hsu, "Design and development of a web based academic advising system", Frontiers in Education Conference, 2001. 31st Annual Publication Year: 2001 , Page(s): S3C - 6-10 vol.3

[17] Dheeraj S. Badde, Anil k. Gupta, and Vinayak K. Patki, "Comparison of Fuzzy Logic and ANFIS for Prediction of Compressive Strength of RMC", IOSR Journal, 2012.

[18] Mehdi Neshat, Ali Adeli , Azra masoumi,and Mehdi sargolzae," A Comparative Study on ANFIS and Fuzzy Expert System Models for Concrete Mix Design", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011.

[17] Mehdi Neshat, Ali Adeli , Azra masoumi, Mehdi sargolzae, "*A Comparative Study on ANFIS and Fuzzy Expert System Models for Concrete Mix Design*", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011

[18] SurinderJassarand Lian Zhao , Zaiyi Liao and Ka Long Ringo Ng, "*Parameter Selection for Training Process of Neuro-fuzzy Systems for average Air Temperature estimation*", Proceedings of the 2009 IEEE International Conference on Mechatronics and Automation, 2012

[19] Arshdeep Kaur, Amrit Kaur, "*Comparison of Mamdani – Type and Sugeno –Type Fuzzy inference systems for Air Conditioning system*", International Journal of soft computing and engineering (IJSCE)", ISSN:2231-2307, Volume -2,Issue-2,May 2012

[20] Gaurav Kumar Nayak, Swathi J Narayanan, Ilango Paramasivam ,"*Development and Comparative Analysis Of Fuzzy Inference Systems for Predicting Customer Buying Behavior*", International Journal of Engineering and Technology (IJET), 2012

[21] Maryam Hassaninia and Hassan Sharafi ," Predicting the pseudo acceleration for near field condition and C type soil using ANFIS model, EJCE,vol18, 2013

