International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(2), pp.329-334
DOI: http://dx.doi.org/10.21172/1.102.53
e-ISSN:2278-621X

PROGRAMMING SKILL CLASSIFICATION MODEL: FFNN APPROACH

Dr. Shruti. S. Jamsandekar[1]

Abstract: -This paper highlights the relevance of a Artificial Neural Network (ANN) learning methodology and attempted to model the task of measuring and classifying computer programming skills of students. ANN model based on Feed Forward Architecture with back propagation learning is devised and trained with two data sets. The model showed a fine capability of classifying the data from educational settings with an accuracy of 90% correct classification in the validation and testing phase, thus the model performance is apposite and acceptable for classification of above stated problem.
Keyword: Educational Data Mining, Classification, Artificial Neural Network

1. INTRODUCTION

Artificial Neural Networks (ANN) is an computational model that has ascended from varied sources, ranging from the simulation of human brain ,mimicking of human reasoning abilities, decision making capabilities to feature distinguishing aptitude (classification) [3].Classification is an technique of grouping datasets into pre-defined classes and has been used extensively in data mining, pattern recognition, rule generation etc[4]. The past and current research activities in classification reveal that the neural networks with its characteristic properties including input-output mapping, learning from past, fault tolerance have made it applicable and favorable substitute to various conventional classification methods[5, 6].

Some of the features of neural networks which has offered an upper edge to machine learning techniques lies in the following theoretical aspects. Neural networks are data driven and self-adaptive as they can organize themselves based on the data provided as the input without any explicit specification of functional or distributional form for the underlying model. It can approximate any function with arbitrary accuracy and are known for universal approximations [8, 9]. Its known for its nonlinear input-output mapping capabilities, which makes them flexible in modeling real world complex relationships.

Also its capability to learn from past, which provides the basis for establishing classification rule [10].Neural network with its inherent parallel architecture can handle parallel input variables and can solve complex classification problems. The application of ANN in these data mining techniques has been pre- dominant with the characteristic advantage of parallel processing, fault tolerance and robustness.

Since any classification procedure seeks for mapping between the class membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important. Neural Network has the ability to extract the patterns and detect multi-dimensional non-linear connections in data. Thus making it more suitable to serve as tool for classification.

Ample literature reports the neural network architecture including single layer feed forward, multi-layer feed forward, recurrent networks, and learning algorithm such as error-correcting learning algorithm (back propagation), Hebbian rule learning, competitive learning rule etc [14-17].Looking into the predictive mining, the most commonly used technique is classification, researchers in [18- 20] have reported the use of ANN having better performance in terms of prediction accuracy and the time utilization. Hongjun Lu et.al in [21] have proposed an approach to discover symbolic classification rules using neural networks. The concise symbolic rules with high accuracy were extracted from a neural network using their approach. The network was first trained to achieve the required accuracy rate, the redundant connections of the network were then removed by a network pruning algorithm

This paper provides a quintessence of feed forward neural networks with back propagation learning employed for classification of student examination data.

2. PROPOSED MODEL

Feed Forward Network with BP Learning Algorithm as Classification Model

The problem is from the educational domain that comprises the task of consulting student examination data, the task involved in classifying programming skill using ANN classification techniques.

3. METHODOLOGY

Step 1: Input/ Output data Identification :

The data set used is the student data set comprising of programming language performance (theoretical and hands on practical). The objective is to classify the students programming skills into different levels.

[1] Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER) Kolhapur.

Inputs – Internal evaluation marks(In), Term end marks(term) and Programming practical marks(prac)
Output – Five classes of skill_level (NA, Level 0, Level 1, Level 2, Level 3)

Step 2 :

Sr. No.	Skill_Level	Description
1	NA	No programming skill assessment done
2	Level 0	Student with very poor programming skills
3	Level 1	Student with poor programming skills
4	Level 2	Student with average programming skills
5	Level 3	Student with good hands on programming and logic

Preprocessing of data in format acceptable to ANN

The data to be used as an input to the network required pre-processing. Many records with partial blank values were assigned value '-1' that is considered as absence during assessment or examination. The 3 Inputs are as one dimensional matrix containing values for each input .

Also during training phase the target data values against which the ANN output class values is compared, had to be encoded in the form of 5X N matrix where 5 is number of output classes and N is the number of data records used for training and testing. A sample target data set is shown in Table 1 for 10 data records, value 1 represent the record classification to a particular output class.

Table 1: 5XN encode target data set

Output Class	Record Target Class									
NA	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	1	1	0	0	0
3	0	0	0	0	0	0	0	0	1	1

Step 3: Network Model created with a feed forward network architecture with back propagation learning algorithm comprising of 2 layers and number of input neurons 3, hidden neurons 20 and output neurons 5.

Step 4 : Configure the network

Training algorithm used is Scaled Conjugate Gradient (trainscg)

Performance function is Mean Square Error (MSE)

Input data randomly divide as 80% training data, 10 % validation data and 10% testing data

bias connection [1;1], weights for input connection: [1; 0], layer connections: [0 0; 1 0] and output connection: [0 1].

Step 5 ANN training Phase – with 2 data sets

Data set 1 with 169 records with 80-10-10 random division of data into training-validation-testing data respectively was experimented. Out of 169 records 17 records are randomly chosen for validation and other 17 records are chosen for testing and rest 135 records for training the network.

Data set 2 with 260 records with 80-10-10 random division of data into training-validation-testing data respectively was experimented. Out of 260 records 26 records are randomly chosen for validation and other 26 records are chosen for testing and rest 208 records for training the network.

The Building of proper classification network required multiple training instances, each of the training instances and the, network built with epochs required to reach the convergence, classification confusion matrix and error rate are recorded.

Step 6 The trained network with best validation performance using Data set 1 and Data set 2 is chosen and taken as classification network to perform further classification test.

Step 7 Classification phase a partially unknown test data set to check for the classification results.

Step 8 The results of classification phase are recorded with number of proper classification and misclassification with performance table. Confusion matrix is plotted.

4. EXPERIMENT WORK AND RESULTS

ANN Creation : The network model is built using feed forward approach with back propagation learning. The neural network is created using MATLAB NNTOOL (Neural Network Tool), following Figure 1 shows the structure of created network and Figure 2 depicts the ANN Model for programming skill classification

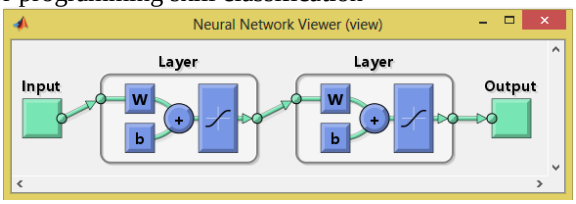

Figure 1: Network Structure

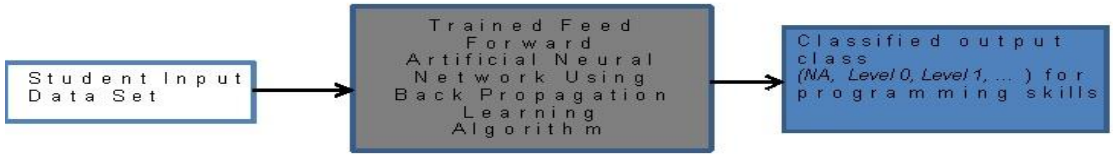

Figure 2: Artificial Neural Network Model for Student Programming Skill Classification

A Feed Forward Neural Network is created with number of layers 2 and number of input neurons 3, hidden neurons 20 and output neurons 5 having bias connection [1;1], weights for input connection: [1; 0], layer connections: [0 0; 1 0] and output connection: [0 1].Training algorithm used is Scaled Conjugate Gradient (trainscg) and performance function is Mean Square Error (MSE) with input data randomly divide as 80% training data, 10 % validation data and 10% testing data.

4.1 ANN Training Phase

ANN training information with data set 1: Data set with 169 records with network structure and architecture as shown in Figure 3 with 80-10-10 random division of data into training-validation-testing data respectively was experimented. Out of 169 records 17 records are randomly chosen for validation and other 17 records are chosen for testing and rest 135 records for training the network. The Building of proper classification network required multiple training instances, each of the training instances and the network built with epoch's required to reach the convergence and error rate are presented in the Table 2

Table 2: Training instances of network

Training Instances No.	No. Epochs	MSE %	Error %
1	56	.022043	0.0577
2	42	.019977	0.0654
3	69	.001943	0.0461
4	27	.027655	0.2000
5	41	.057659	0.2346
6	56	.052805	0.2269
7	63	0.10236	0.3846
8	24	.030056	0.1962
9	44	.096597	0.2846
10	80	.012931	0.0423

Looking into class wise, proper classification is elaborated in the Table 3 with the percentage of correct and incorrect classifications. The best trained network is shown in training instance number 10 with 4.2 % error of total trained network with random division of data set as 80-10-10 for training, validation and testing, respectively.

Table 3 : Network training instances

Training Instance No.	Output Class- % classification					Network -% error
	1	2	3	4	5	
1	100	91.7	97.9	92.7	92.3	5.8
2	100	91.7	89.6	92.7	95.4	6.5
3	100	87.5	95.8	97.9	92.3	4.6
4	0	91.7	95.8	85.4	89.2	20.0
5	100	87.5	0	93.8	93.8	23.5
6	100	87.5	0	95.8	93.8	22.7
7	100	87.5	0	97.9	100.0	38.5
8	0	91.7	91.7	89.6	87.7	19.6
9	100	87.5	89.6	99.0	0	28.5
10	100	91.7	95.8	96.9	93.8	4.2

The network training instance number 10 is chosen and taken as classification network to perform further test. Best validation performance of the network is 0.012931 MSE at 74th epoch, the performance plot is shown in Figure 3 and the confusion plot of the network is presented in Figure 4.

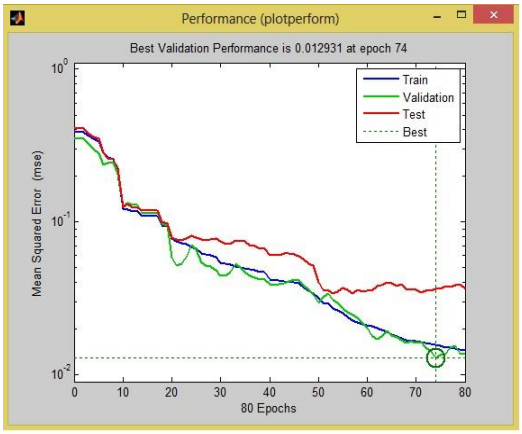

Figure 3: performance plot of network

Figure 4: confusion matrix for classification twork

ANN training information with Data Set 2: Data set with 260 records with same network structure and architecture as above with 80-10-10 random division of data into training-validation-testing data, respectively. Out of 260 records 26 records are randomly chosen for validation and other 26 records are chosen for testing and rest 208 records training the network. The training instances of the network built with epoch's required to reach the convergence and error rate are presented in the Table 5.

Table 5: Training instances of network

Training Instances No.	Number of Epochs	MSE %	Error %
1	19	0.08944	0.2154
2	30	0.081587	0.2769
3	91	0.045091	0.1615
4	23	0.11472	0.2000
5	47	0.083308	0.2231
6	44	0.046852	0.0885
7	72	0.017015	0.0308
8	61	0.0090062	0.0500
9	21	0.089907	0.4500
10	31	0.01519	0.1885

Table 6 elaborates the percentage of correct classifications of each output class for different training instances of the network. The network training instance number 7 is chosen and taken as classification network to perform further test. Best validation performance of the network is .017015 MSE at 66th epoch. The performance plot is shown in Figure 5 and the confusion plot of the network is presented in Figure 6

Table 6: Network training instances

Training Instance No.	Output Class- % classification					Network - % error
	1	2	3	4	5	
1	100	0	87.5	85.4	81.5	21.5
2	100	79.2	95.8	100.0	0	27.9
3	0	100	97.9	91.7	90.8	16.2
4	100	75.0	89.6	97.6	0	30.0
5	0	33.3	95.8	90.6	93.8	77.7
6	100	83.3	95.8	90.6	87.7	8.8
7	100	91.7	100.0	95.8	96.9	3.1

8	100	87.5	95.8	95.8	93.8	5.0
9	100	83.3	0	100.0	0	45.0
10	0	95.8	89.6	90.6	89.2	18.8

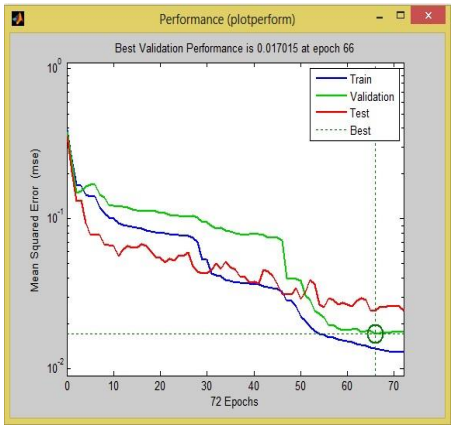

Figure 5: performance plot of network

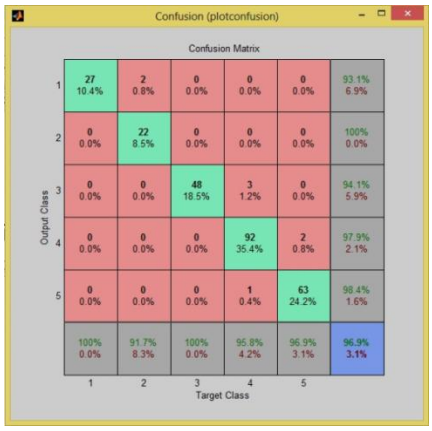

Figure 6 : confusion matrix for classification

4.2 ANN Testing and Classification

In the classification phase we have employed a partially unknown test data set to check for the classification results. Figure 7. shows the confusion matrix test data set on both the Neural Networks created from two different training data set, coincidentally both the created artificial neural network system has classified 27 cases correctly out of 30 given input student records. They have given 90% correct classification and 10% misclassification result. Testing dataset contains four type-NA classes, four type-0 classes, seven type-1 classes, eight type-2 classes and seven type-3 classes From the confusion matrix shown in Figure 8 and performance shown in Table 7 we observe that out of given Seven type-3 class cases it classified five cases correctly as type-3 and remaining two cases classified wrongly as type-2. For the Eight type-2 class cases it classified seven cases correctly and one case wrongly as type-3.

Table 7: Performance table for test data on network

class	False negative	False Positive	True Positive
NA	0	0	1
0	0	0	1
1	0	0	1
2	0.1250	0.2500	0.8750
3	0.2857	0.1429	0.7143

Figure 7: Test data confusion matrix

5. CONCLUSION

The Classification model based on ANN approach using feed forward architecture with back propagation learning was devised and trained with two data sets. The dataset-I with 169 records has best validation performance of the network with 0.017015 MSE and dataset- II with 260 records has best validation performance of the network with 0.012931 MSE. The proposed ANN model showed a fine capability of classifying the data from educational settings with 90% of correct classification and 10% misclassification helping the model to be well suited as a classifier for the given domain problem.

6. REFERENCES :

[1] Arifs. AI-Hammadi and R. H. Milne, "A Neuro-Fuzzy Approach for Student Performance Modeling", IEEE conference proceedings ICECS-2003.

[2] Regina Stathacopoulou, Maria Samarakou, Maria Grigoriadou, George, D.Magoulas, "A Neuro-Fuzzy Approach to Detect Student's Motivation", Proc. of the IEEE Int. Conf. on Advanced Learning Technologies (ICALT'04)

[3] Ben Krose,Patrick van der Smagt, An introduction to Neural Network, The University of Amsterdam, Eighth Edition,1996.

[4] Jiawei Han and MichelineKamber, Data Mining: Concepts and Techniques,2nd ed. Morgan Kaufmann Publisher.

[5] David Reby,SovanLek, IoannisDimopoulos, Jean Joachim, Jacques Lauga, Stephane Aulagnier, "Artificial neural networks as a classification method in thebehavioral sciences", Elsevier Publication, Behavioral Processes, Vol.40, pp.35–43, 1997.

[6] R. Sathya, Annamma Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification", International Journal of Advanced Research in Artificial Intelligence(IJARAI),Vol. 2(2), 2013

[7] Saravanan K and S. Sasithra, "Review On Classification Based On Artificial Neural Networks", International Journal of Ambient Systems and Applications (IJASA), Vol.2(4), December 2014.

[8] KurthHornik, "Degree of approximation Results for Feed Forward Networks Approximation unknown mappings and their derivatives", NeuroCLTO technical report series, 1995.

[9] Tiapingchen, Hong Chen, Leu, "Approximation Capability in C(R)n in Multilayer Feed Forward Network and Related Problems", IEEE transactions on Neural Network.

[10] Ping Chang and Jeng-Shong Shih, "The Application of Back Propagation Neural Network of Multi-channel Piezoelectric Quartz Crystal Sensor for Mixed Organic Vapours", Tamkang Journal of Science and Engineering, Vol. 5(4), pp. 209-217, 2002.

[11] R. Rojas, "The Backpropagation Algorithm", Chapter 7 Neural Networks, Springer-Verlag, Berlin, 1996.

[12] Dian Pratiwi, Diaz D. Santika, and Bens Pardamean, "An Application of Backpropagation Artificial Neural Network Method for Measuring the Severity of Osteoarthritis", International Journal of Engineering and Technology IJET-IJENS, Vol. 11(03), pp.102-105,2011.

[13] Zhen-Guo Che1, Tzu-An Chiang2 and Zhen-Hua Che3, "Feed-Forward Neural Networks Training: A Comparison Between Genetic Algorithm and Back-Propagation Learning Algorithm", International Journal of Innovative Computing, Information and Control, Vol. 7(10), pp. 5839-5850, October 2011.

[14] Detlef Nauk,"Neuro Fuzzy System Review and Prospective", Proc.of Fifth European Congress On Intelligent Techniques And Soft Computing,pp.1044-1053,Aachen Sept 1997.

[15] Robert Fuller, Neural Fuzzy Systems, Abo Akademi University, ISBN 951-650-624-0, ISSN 0358-5654, 1995.

[16] R. Fuller, "Introduction to Neuro-Fuzzy Systems", Series: Advances in Intelligent and Soft Computing, Vol. 2, pp. 289, 2000.

[17] Sushmita Mitra, Sankar K. Pal, Pabitra Mitra, "Data Mining In Soft Computing Framework: A Survey", IEEE transactions on neural networks, Vol. 13(1), January 2002.

[18] K Jaimin, N. Undaviaa, P.M.Doliab, Atul Patela, "Comparison of Classification Algorithms to Predict Comparison of Decision Tree Classification Algorithm to Predict Student's Post Graduation Degree in Weka Environment ", International Journal of Innovative and Emerging Research in Engineering, Vol. 1(2), pp. 17-22, 2014.

[19] Armender kumar, Artificial neural network for data mining, I.A.S.R.I Library, pp. 157-170.

[20] Dr.Yashpal Singh, Alok Singh Chauhan, "Neural Networks In Data Mining", Journal Of Theoretical And Applied Information Technology,Vol.5(6) , pp.37-42, 2005.

[21] Hongjun Lu, Member, IEEE Computer Society, Rudy Setiono, and Huan Liu, Member "Effective Data Mining Using Neural Networks",IEEE transactions on knowledge and data engineering, Vol. 8(6), pp. 957- 961,1996.