

Design of an Intelligent Network for Classification of Data for Recruitment Using Neuro-Fuzzy Network

¹Ritesh Khedekar
¹M.E. Student,

²Arvind Upadhyay
²Associate Professor,

¹M.E Student, Dept. of CSE, Institute of Engineering & Science(IES),IPS Academy, Indore, MP, India

²Associate Professor, Dept. of CSE, Institute of Engineering & Science(IES),IPS Academy, Indore, MP, India

Abstract

Student recruiting is one of the biggest issues in the industries and institutions that have direct impact on budget planning and education management. So selection of the right candidate for a particular organization is a typical task for HR manager. It consumes lot of time, efforts and investment for an organization. Initially the organization goes for the resume filtering or selection process for selecting the right candidate from many resumes. Candidates are filtered on some criteria. The main objective of this research is resume filtering or selecting candidate using (NeuroFuzzy) or Proposed method. The objective has been achieved by data collection from the engineering institute to support the research. The Data mining clustering [11][4] and classification techniques[3] such as fuzzy c-mean, artificial back propagation neural network(BPNN) and proposed method (Neuro-Fuzzy) have been applied to discover unknown knowledge. In experiment phase, we used selected classifier algorithm in order to propose the suitable classifier from student dataset. From comparative study it has been observed that proposed method (Neuro-Fuzzy) performs 4-5% better results than fuzzy c-mean and artificial neural network (BPNN) [6][10].

Index terms – Data mining, Fuzzy C-mean, Artificial Neural network, Clustering, Classifiers, BPNN

I. INTRODUCTION

The recruitment process is one of the important functions of the HR department in any organization and resume selection is the first step towards creating the good and effective staff for the organization. Recruitment of the right candidate is also important for IT companies as well as for the engineering institute. The mechanism used by both organization may differ but commonly attributes include education qualification, experience, marks obtained in test, communication skill etc. One of the major problems associated to recruitment is resume filtering as it requires lot of efforts to analyze the profile of all the candidates as per the need of organization. A single vacancy can have lot of candidate in today scenario. So the ratio of selecting a single or few candidates among thousands of applied candidate will require a lot of time and man-power. Reducing this ratio will help the organization to save time and money.

In this research, a framework is proposed in order help organization manage recruitment effectively using data mining techniques. Data mining is the process of extracting knowledge from large amount of data using artificial intelligence, machine learning etc. It can make model and discover the unknown knowledge. We have proposed the model for engineering students based on fuzzy logic and neural network. In this paper results of the experiments were conducted using clustering and classification techniques like Fuzzy c-mean and Back propagation neural network algorithm.

II. Literature Survey or Related works

N. Sivaram and K. Ramar [1] have presented research by using data mining techniques[5] to categorize and classify the features of data analysis for candidates in the IT companies to find patterns of those who are selected or unselected to work. They compare to find the suitable algorithm and the results showed that the clustering algorithms are not appropriate to the problem. It should be the decision tree classification algorithm such as ID3, C4.5 and CART with C4.5 showing the highest accuracy.

O.S Akinola and B.O. Akinkunmi [2] have presented data mining technique to categorize and classify the feature of data analysis for predicting computer programming efficiency of computer science undergraduate students. This study employed the use of Artificial Neural network data mining tool to predict the performance of students in computer programming. Results from the study shows that apriori knowledge of Physics and Mathematics are essential in order for a student to excel in computer programming.

H. Jantan et al. [3], have presented data mining technique to classify talent knowledge acquisition process in Human Resource by using classification techniques. The challenge for HR is to ensure the right person assign to the right job at the right time. The purpose of this study is to identify potential data mining technique for selecting the right talent. The first technique chosen is neural network and used as pattern classification. The second technique is decision tree known as 'divide and conquer' and third technique is nearest neighbor which is based on distance metric. In experimental phase, they used C4.5 and Random forest for decision tree; Multilayer Perceptron (MLP) and Radial Basic Function Network (RBFN) for neural network; and k-star for the nearest neighbor technique. Two dataset were used. The first dataset was concentrating on academic talent and second dataset was focusing on employee performance evaluation through yearly performance. The performance attributes were identified from yearly performance appraisal records, previous expert knowledge and expertise records. The DM tools used were WEKA and ROSETTA toolkit. The experiment

was focused on the accuracy of the classifier in order to identify the suitable classifier algorithm for the talent datasets. The accuracy of classifiers was based on the percentage of test set samples which were correctly classified. As a result, the C4.5 classifier algorithm from decision tree family is recommended as a suitable classifier for the datasets.

2.1 Data Mining Process

We prepared data consisting of the following procedures:

Data preparation: The data set used in this study was obtained from an engineering college. It includes various attributes such as 10th, 12th, UG, PG courses done and their percentage.

Data Selection: In this step only those records and field were selected which are required for data mining.

Data Cleaning: This step is needed to remove the errors due to incomplete information, wrong feeding and cutting off the attribute that is not important.

Data Transformation: Data must be in proper format to meet the requirement of data mining application. We used Ms-Excel format whose normalize values are supplied to different algorithms.

The results of data preparation yields dataset1 of 100 records and dataset2 of 200 records and 7 attributes: High School Marks, High Secondary marks, undergraduate, postgraduate marks, GATE, industry experience and institute experience.

2.2 Data mining techniques

The machine learning is a subfield of artificial intelligence which is concerned with the development of algorithms and techniques that allow computers to learn the ability of identifying, classifying, and predicting of data set. The Fuzzy c-mean and back propagation neural networks are widely applied methods in machine learning.

2.2.1 Fuzzy C-mean Technique

Fuzzy c-mean clustering (FCM)[18] is a method of clustering which allows one point to belong to one or more clusters. The method was developed by Dunn in 1973 and improved by Bezdek in 1981 and it is frequently used in pattern recognition. The FCM algorithm attempts to partition a finite collection of

points into a collection of C-fuzzy clusters with respect to some given criteria. Thus the points on the edge of the cluster to a lesser degree than points in the center of cluster.

FCM algorithm[13] attempts to partition a finite collection of n elements $X=\{X_1, X_2, X_3, \dots, X_i\}$ into a Collection of C fuzzy clusters. Given a finite set of data, the algorithm returns a list of clusters centers $C=\{C_1, C_2, C_3, \dots, C_j\}$.

With Fuzzy c-mean, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster. The algorithm begins by choosing the number of clusters and fuzzification parameter. Center for all the clusters are chosen randomly. The algorithm continues to update the center of the clusters till the value stabilizes.

Algorithm for Fuzzy c-means clustering [15]

Let $X=\{x_1, x_2, x_3, \dots, x_i\}$ be the set of data points and $V=\{v_1, v_2, v_3, \dots, v_j\}$ be the set of centers.

1. Randomly select 'c' cluster center.
2. Calculate the fuzzy membership using

$$\mu_{ij} = 1 / \sum_{k=1}^n (d_{ij} / d_{ik})^{2/(m-1)}$$

3. Compute the fuzzy centers using

$$v_j = \left(\sum_{i=1}^n (\mu_{ij})^m x_i \right) / \sum_{i=1}^n (\mu_{ij})^m$$

for all $j=1, 2, 3, \dots, c$

where

'n' is the number of data points

'v_j' represents the jth cluster center .

'm' is the fuzziness index m.

'μ_{ij}' represents the membership of ith data to jth cluster

'd_{ij}' represents the Euclidean distance between ith data and jth cluster center.

'c' represents the number of cluster center

4. Repeat step 2 and 3 until the minimum value of J is achieved

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|X_i - V_j\|^2$$

Where $\|x_i - v_j\|$ is the Euclidean distance between ith data and jth cluster center.

2.2.2 Back propagation Neural Network Technique (BPNN)

The back propagation algorithm[19] was developed by Paul Werbos in 1974 and rediscovered independently by Rumelhart and Parker. It is a method of teaching artificial networks how to perform a given task. It is used in layered feed forward ANN. The artificial neurons are organized in layers and send their signals "forward" and then the errors are propagated backwards. It used supervised learning, which means that we provide the algorithm with examples of inputs and outputs we want the network to compute and then the error (difference between actual and expected results) are calculated. The idea of back propagation is to reduce this error. Error data at the output layer is "back propagated" to earlier ones, allowing incoming weights to these layers to be updated. To update the weights, one must calculate an error. At the hidden layers, however, there is no direct observation of the error; hence, some other technique must be used. To calculate an error at the hidden layers that will cause minimization of the output error, as this is the ultimate goal.

Steps:

1. Present a training sample to the neural network.
2. Compare the network's output to the desired output from that sample. Calculate the error in each output neuron.
3. for each neuron, calculate what the output should have been, and a scaling factor, how much lower or higher the output must be adjusted to match the desired output. This is the local error.
4. Adjust the weights.

BPNN Algorithm:

1. Initialize the weights in the network (often randomly)
2. Repeat
 - * for each example e in the training set do
 1. O = neural-net-output (network, e);
Forward pass
 2. T = teacher output for e
 3. Calculate error (T - O) at the output units
 4. Compute delta_wi for all weights from hidden layer to output layer;
Backward pass
 5. Compute delta_wi for all weights from input layer to hidden layer;

- Backward pass continued
 6. Update the weights in the network
 * end
3. until all examples classified correctly or stopping criterion satisfied
 4. Return (network)

III. PROPOSED MODEL

The problem domain of selecting the right candidate from a large dataset is quit complex and the tedious task. Because of the inconsistency in the quality of the students produced by different universities and the type of skill set they acquire during their program, selecting the right candidate becomes a difficult task. The design of the system requires domain experts to obtain the required information to solve the problem, knowledge extraction[3] was made with the collected information and a knowledge base is built. The proposed model is as follows:

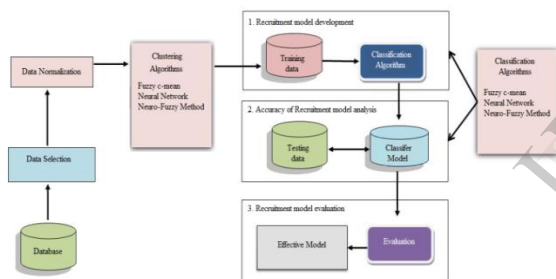


Figure 1. Proposed model

However many researches have been done to analyse the system and concluded that a pattern exist among the candidates selected for an organization. Data mining techniques such as Fuzzy C-mean , K-means and Neural network[20] have been employed. The percentage accuracy varies according to the input provided. But it has been observed that fuzzy C-mean and Back propagation network accuracy can be further improved if we combine the two techniques i.e if we use a mixed method which is a combination of Neuro-fuzzy [12] system that can be used to model the intelligent system for learning. The percentage accuracy can be much improved if we go for neuro-fuzzy technique rather than fuzzy c-means and neural network individually.

Proposed Algorithm

- Construct the database as per domain experts. //Construction of knowledgebase
 Target= selected candidate as per domain expert.
- Upload the database at specific path
 //Selection of data set and preprocessing
 File name=getfile(*.*)
 W= read(filename)
 Data=w(relevant attributes)
 Plot(data values)
- Pick the normalization values
 $m=[100 \ 100 \ 100 \ 100 \ 4 \ 4 \ 1]$ // Data normalization
 For $i=1$ to size(Data)
 $data1 = data/m$;
 Plot(normalize values)
 end for .
- If Algorithm=fuzzy //Apply fuzzy Algorithm
 Fuzzy=genfis(data1)
 Out1=fuzzy
 Accuracy=100*Sum(out1=target)
 Plot (fuzzy values)
 else
 If Algorithm= BPNN //Apply Back propagation Neural network
 Neuralnet=gennewnet(data1)
 Out2=Neuralnet
 Accuracy=100*Sum(out2=target)
 Plot(neuralnet values)
 Else
 If Algorithm=neuro-fuzzy // Apply proposed method (neuro-fuzzy)
 Out=Out1+Out2 //Neuro-fuzzy=fuzzy+neuralnet
 If Out>=1
 Accuracy=100* Sum(out==target) // Compare out with domain expert output.

5. Compare the result of number of selected candidate of and choose the algorithm with highest accuracy.

The application of data mining based on neural networks can be divided into different stages: Knowledge acquisition, data preparation, and modeling and knowledge discovery. The data sets and the input attributes are determined through knowledge of an engineering college. The mining process begins with the step to gather knowledge from the domain experts. Knowledge acquisition includes initiation, collection, analysis, modeling and validation of knowledge. The knowledge acquired is used along with the recruitment database maintained to form the dataset for experiment. The data is preprocessed to remove missing and inconsistent data and hence improve the quality of data and make it fit for data mining task. The data preprocessing includes:

Data Selection: The selection of data with important feature that is unique to a particular group from a database.

Data Cleaning: This step removes errors by removing incomplete information and cutting off the attribute that is not important.

Data Transformation: Data must be in a format to meet the requirements of the data mining application. Using normalization we provide the numerical data type to the neural network which only accepts the data between 0 and 1 as input.

Clustering techniques[4] were applied for the data and the constructed models were reviewed and evaluated using classifiers. The models (Fuzzy c-mean, back propagation and Neuro-fuzzy) were evaluated using accuracy as the criteria to access the performance of the different techniques. Accuracy is determined as the ratio of records correctly classified during testing to the total number of records tested. There are two aspects to view the Neuro-fuzzy system:

1. One is to enable Neural Network with fuzzy capabilities, thereby increasing the flexibility to adapt to uncertain environments.
2. To apply neuron learning capabilities to fuzzy system to make the fuzzy system more adaptive to changing environment.

The modified neural network[17][[9][7] is nothing but combination of Fuzzy C-mean and Back

propagation neural network. The algorithm includes the fuzzy concept of partial membership and back propagation algorithm that reduce the error by differencing the actual and expected results.

IV. EXPERIMENTAL RESULTS

The system proposed in this paper has been implemented and evaluated on two different datasets. This section presents the details of the data sets, test results and comparison of them. The metric used to evaluate the clustering and classification algorithm is the accuracy. The first dataset1 includes the details of the engineering students that consist of 100 records, the second dataset2 includes 200 records. The algorithms were trained with records of one dataset and tested with the records in the other dataset. The dataset contains input table with the desired output. The data selection on these data set1 is applied. After data preprocessing apply the different techniques. After training the dataset1, it is self tested using the classifier. From the dataset it is observed that the dataset consist of more than 75% of records to be in selection category. Hence algorithms were very excellent in recognizing the dataset. Clustering and classification techniques were applied with Matlab 7.8.0.347 and the accuracy of the clustering techniques is depicted in table .

1. Trained with dataset1(100 records) and tested with dataset2 (200 records)

Table 1

Algorithm	Accuracy
Fuzzy c-mean	55%
Neuro-fuzzy(proposed)	77%
Neuro-fuzzy	78%

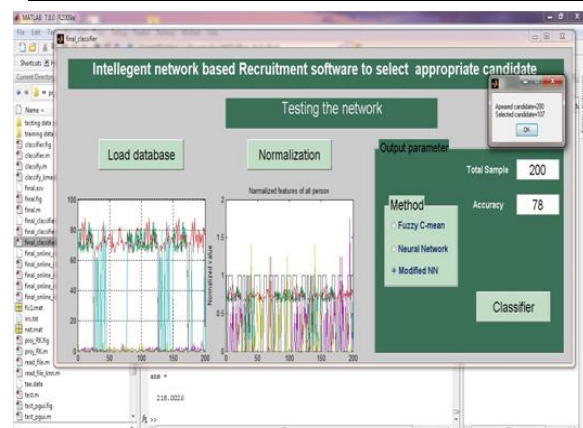


Figure2 Showing the result on Matlab for 200 records.

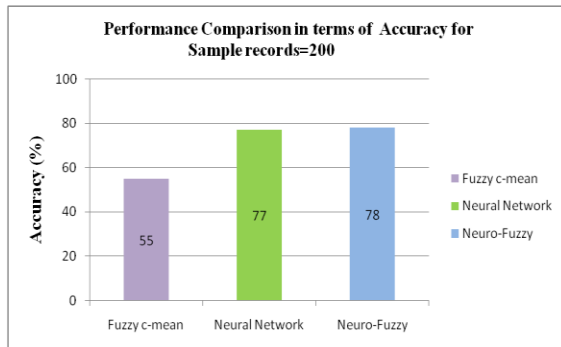


Figure 3 Performance comparisons Accuracy of the 3 techniques testing result of 200 records.

2. Trained with dataset1 (100 records) and tested with dataset3 (400 records).

Table 2

Algorithm	Accuracy
Fuzzy c-mean	54.5%
BPNN	77.5%
Neuro-fuzzy(Proposed method)	80.5%

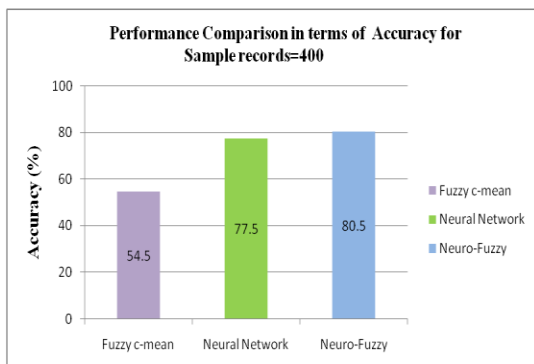
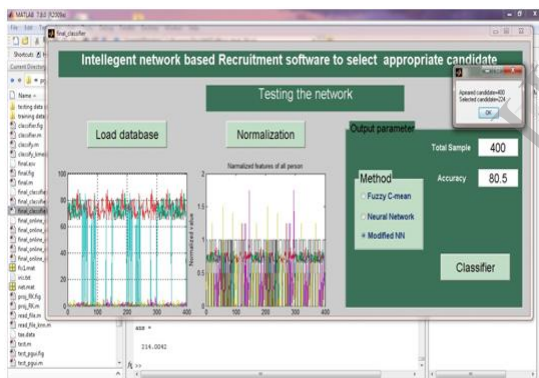


Figure 4. Performance comparison Accuracy of the three techniques with testing results of 400 records

3. Trained with dataset1 (100 records) and tested with dataset4 (800 records)

Table 3

Algorithm	Selection Percentage
Fuzzy c-mean	60.75%
BPNN	76.75%
Neuro-fuzzy	80.37%

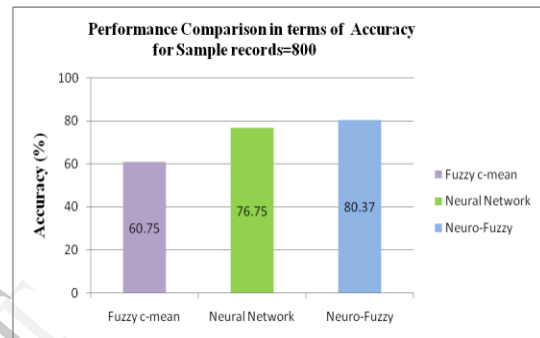


Figure 5 Performance comparisons Accuracy of the 3 techniques testing result of 800 records.

From comparative analysis it is observed that the Accuracy of Modified NN (Neuro-Fuzzy) is 75% or above and in every dataset whether it is data of 200 records, 400 records and 800 records 75% of students where in selection list.

V.CONCLUSION

This research aims to study the selection of students who have completed engineering degree. Based on selection criteria by expert we have to select appropriate candidate. In this paper we present a neuro-fuzzy based approach to data mining. The approach consists of different phases.

1. Constructing and training a network to correctly classify records in the given training data set to required accuracy.
2. Comparing the algorithm fuzzy c-mean, BPNN with proposed method (neuro-fuzzy) to compare the results and conclude which has better accuracy and hence which model is best suited for a particular dataset.

A set of experiments was conducted to test the proposed approach. The results indicate that, using the proposed approach, useful data can be discovered from the given data sets. In future we can apply the technique to select the deserving candidates for an organization. Hence the use of neural network and fuzzy system in data mining can be used in research when large mass of data sets need to assimilate relationship between a large number of variables. The main purpose of this study is to improve the performance of BPNN by introducing a model that can enhance the learning capability and accuracy of the results. The experimental results show that the proposed Neuro-fuzzy model can achieve better performance compared to the standard BPNN model.

VI. REFERENCE

- [1] N.Sivaram and K.Ramar , “Applicability of Clustering and Classification Algorithms for Recruitment Data Mining”, International Journal of Computer Applications(0975-8887) Volume 4-No.5, July 2010
- [2] O.S. Akinola, B.O. Akinkunmi & T.S. Alo, "A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students," IEEE, African Journal of Computing & ICT, Vol-5, No.1 pp43-52,January 2012.
- [3] Hamidah Jantan, Abdul Razak Hamdan and Zulaiha Ali Othman, "Talent Knowledge Acquisition using Data Mining Classification Techniques," 3rd conference on data mining and optimization, Selangor, Malaysia, IEEE, 28-29 June 2011.
- [4] Osama Abu Abbas, "Comparison between Data Clustering Algorithms", International Arab Journal of Information Technology, Vol.5, No.3, July 2008
- [5] L Wang and T.Z. Sui,, "Data Mining Technology Based on Neural Network in the Engineering", School of Mechanical Engineering and Automation ortheastern University IEEE 2007.
- [6] Dr. Yaspal Singh and Alok Singh Chauhan “Neural networks in Data Mining”, Journal of Theoretical and Applied Information Technology JATIT, 2009.
- [7] Mobarakol Islam, Arifur Rahaman, Md. Mehedi Hasan, Md. Shahjahan,” An Efficient Neural Network Training Algorithm with Maximized Gradient Function and Modulated Chaos”, Fourth International Symposium on Computational Intelligence and Design, IEEE,2011.
- [8] Bing Gong, “A Novel Learning Algorithm of Back-propagation Neural Network”, IITA International Conference on Control, Automation and Systems Engineering, IEEE,2009.
- [9] Asrul Adam, Lim Chun Chew and Junzo Watada, “A Modified Artificial Neural Network Learning Algorithm for Imbalanced Data Set Problem”, Second International Conference on Computational Intelligence, communication system and Networks, IEEE, 2010.
- [10] S.N. Sivanandam and S.N. Deepa, “Principle of Soft Computing”, Wiley India Pvt Ltd, 2011.
- [11] Jiawei Han, Micheline Kamber. 2006. Data Mining Concepts and Techniques”, Second Edition Morgan Kaufmann Publishers, San Francisco.
- [12] Chung-Kwan Shin, Ui Tak Yun, Huy Kang Kim, and Sang chan Park, “A Hybrid Approach of Neural Network and Memory Based Learning to Data Mining”. IEEE transactions on Neural network, Vol.11, No.3, pp.637-646,2000
- [13] Shruti S Jamsandekar , R R Mudkholkar , “Performance evaluation by fuzzy inference technique” ,International Journal of Soft Computing and Engineering (IJSCE), ISSN:2231-2307, May2013
- [14] Bharadwaj, B.K. and Pal. S. “Data mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security(IJCSIS), Vol.9 No.4, pp.136-140.
- [15] C. F. Chien and L. F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry," *Expert Systems and Applications*, vol. 34, pp. 380-290, 2008.
- [16] K. K. Chen, M. Y. Chen, H. J. Wu, and Y. L. Lee, "Constructing a Web-based Employee Training Expert System with Data Mining Approach," *Paper in The 9th IEEE International Conference on Ecommerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007)*, 2007.
- [17] M. J. Huang, Y. L. Tsou, and S. C. Lee, "Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge," *Knowledge-Based Systems*, vol. 19, pp. 396-403, 2006.
- [18] Sushmita Mitra, Sankar K. Pal, Pabitra Mitra, “Data Mining in Soft Computing Framework: A Survey”, IEEE Transactions on Neural Networks, vol. 13, No. 1, pp. 3-14, 2002.
- [19] S. Rajasekaran and G.A. Vijayalakshmi Pai ,” Neural Networks, Fuzzy Logic, and Genetic Algorithms”, PHI Learning Private Limited,2010