A NEURO FUZZY EXPERT SYSTEM FOR HEART DISEASE DIAGNOSIS

E.P.Ephzibah[1], Dr. V. Sundarapandian[2]

[1] School of Information Technology and Engineering,
VIT University, Vellore, TamilNadu, India,

ep.ephzibah@vit.ac.in.

[2]Research and Development Centre,
Vel Tech Dr. RR & Dr. SR Technical University,
Chennai, Tamil Nadu, INDIA.

sundarvtu@gmail.com

ABSTRACT

Heart disease in India is one of the major causes of death. This disease is common not only in old and middle aged people but also in young people. It is caused due to improper diet habits. The proposed system finds a solution to diagnose the disease using some of the evolutionary computing techniques like genetic algorithm, fuzzy rule based learning and neural networks. The neuro fuzzy classification of the disease with the help of genetic algorithms for feature selection is the frame work of the proposed system.

KEYWORDS

Fuzzy rule based expert system, genetic algorithm, heart disease diagnosis, neural networks.

1. INTRODUCTION

According to a survey conducted by India Today Heart diseases in India have emerged as number one killer in both urban and rural areas of the country. It will be of greater value if the disease is diagnosed in its early stage. Correct diagnosis of the disease will decrease the death rate due to heart failures. Many clinical tests are being done to find the presence of the disease. According to the proposed system fewer numbers of tests can effectively help in diagnosing the disease. The proposed system uses the   genetic algorithms for feature   selection which reduces the features so that diagnosis can be done with limited number of tests. Having obtained the features, they are taken to the classifier for prediction with the help of fuzzy rules. The classifier used in this work is NN Classifier.

2. RELATED WORK

In recent years there were many methods proposed for the generation of fuzzy rules. Hong and Lee (1996) presented a method for inducing fuzzy rules and membership functions from training instances to deal with the Iris data classification problem. Hong and Lee (1999) discussed the effect of merging order on performance of fuzzy rules induction. Hong and Chen (1999) presented a method to construct membership functions and generate fuzzy rules from training instances by finding relevant attributes and membership functions to deal with the Iris data classification problem. A method by Castro et al. (1999) generates fuzzy rules from training data to deal with the Iris data classification problem. A method proposed by Chang and Chen (2001) generates weighted fuzzy rules to deal with the Iris data classification problem. Chen and Chen (2002) presented a method based on genetic algorithms to construct membership functions and fuzzy rules to deal with the Iris data classification problem. Chen and Chang (2005) presented a method to construct membership functions and generate weighted fuzzy rules from training instances. Chen and Tsai (2005) presented a method to generate fuzzy rules from training instances to deal with the Iris data classification problem. Chen and Fang (2005a) presented a method for handling the Iris data classification problem. Chen and Fang (2005b) presented a method to deal with fuzzy classification problems by tuning membership functions for fuzzy classification systems. Chen et al. (2006) presented a method for generating weighted fuzzy rules from training data for dealing with the Iris data classification problem.

3. GENETIC ALGORITHMS

Genetic Algorithm is an optimization technique that yields a solution that fits properly into the objective function. It is a form of generate and test paradigm. It works by creating an initial population of N possible solutions in the form of candidate chromosomes or individuals. The individual chromosomes are the representation of the final solution. The objective function used pays a very important role in measuring the goodness of fit of the individual chromosome. Better the fit, the closer the individual to the target. The evaluation process continues till all the individual chromosomes are evaluated. The termination condition is the stopping criteria. The termination condition depends on any of the following:

    a. Maximum number of generations,
    b. Maximum amount of computing time,
    c. Chromosomes having the fitness value that satisfy the objective function etc…

The genetic operators like the selection, crossover and mutation can influence in finding the optimal and feasible solution to the problem. In our work we have used the Roulette wheel selection and the values for crossover and mutation are .5 and .08 respectively.

4. FUZZY RULE GENERATION STEPS

The review on the fuzzy sets is given below:
Zadeh introduced the term *fuzzy logic* in his seminal work "Fuzzy sets," which described the mathematics of fuzzy set theory (1965). Plato laid the foundation for what would become fuzzy logic, indicating that there was a third region beyond True and False. It was Lukasiewicz who first proposed a systematic alternative to the bivalued logic of Aristotle. The third value

Lukasiewicz proposed can be best translated as "possible," and he assigned it a numeric value between True and False. Later he explored four-valued logic and five-valued logic, and then he declared that, in principle, there was nothing to prevent the derivation of infinite-valued logic. FL provides the opportunity for modeling conditions that are inherently imprecisely defined. Fuzzy techniques in the form of approximate reasoning provide decision support and expert systems with powerful reasoning capabilities. The permissiveness of fuzziness in the human thought process suggests that much of the logic behind thought processing is not traditional twovalued logic or even multivalued logic, but logic with fuzzy truths, fuzzy connectiveness, and fuzzy rules of inference. A fuzzy set is an extension of a crisp set. Crisp sets allow only full membership or no membership at all, whereas fuzzy sets allow partial membership.
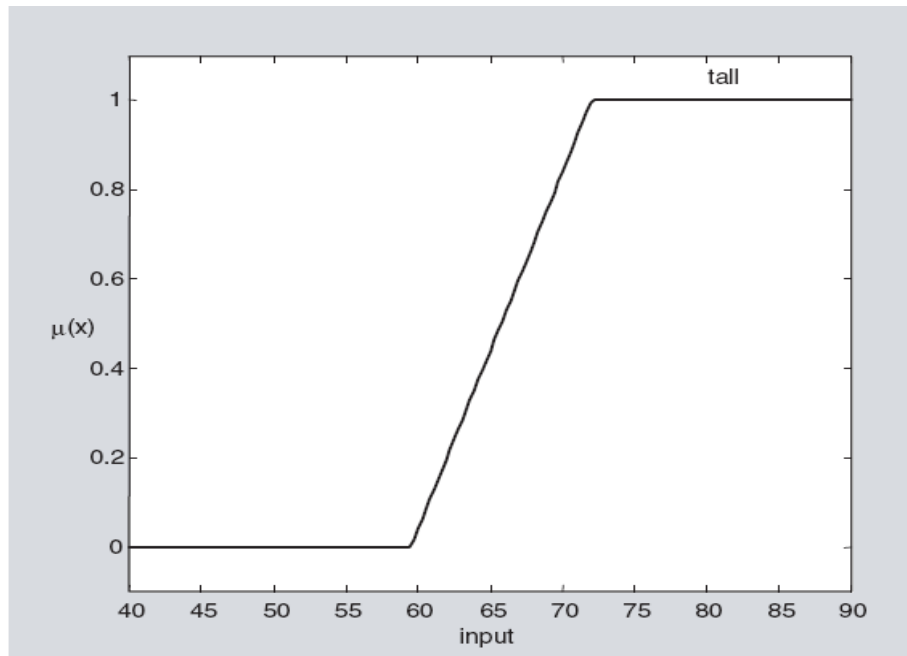

Figure1: An example of a fuzzy membership function.

Fuzzy set operations are analogous to crisp set operations. The important thing in defining fuzzy set logical operators is that if we keep fuzzy values to the extremes 1 (True) or 0 (False), the standard logical operations should hold.

According to S.-M.Chen, F.M.Tsai, the following steps are used to frame the fuzzy rules for the proposed system.

Steps involved:

I: Get the highest and lowest values of the attributes.
II: Find out the degree of entropy of each attribute.
III: Select the attributes whose degree of entropy is more than the   threshold value.
IV: Choose the attribute X found in step III that has the highest degree of entropy, and find the highest attribute values and the lowest attribute values of the attribute X of each species based on the results obtained in step I. Arrange these attribute values in an ascending sequence. Subtract

0.05 cm from the attribute values that are at odd positions, and add 0.05 cm to the attribute values that are at even positions. Then, once again arrange these values in a sequence again. These are the crossover points corresponding to adjacent membership functions of the linguistic terms of the X.

V: Find out the statistical distribution of the attribute values of X of the training instances falling in each interval. The rules are generated from the training instances. If most of the training instances fall in an interval belonging to the species Z1, and the corresponding linguistic term of this interval is Y1, then generate the following fuzzy rule:

IF X is Y 1 THEN the flower is Z1.

VI: Calculate the classification accuracy rate of the generated fuzzy rules. If the classification accuracy rate of the generated rules is equal to or larger than the classification threshold in the range [0,1], then Stop. Otherwise, go to Step VII.   If it is not equal, then it means that this training instance is classified incorrectly.

The classification accuracy rate of the generated fuzzy rules is defined as given below:

$$\text{Classification Accuracy Rate} = \frac{\text{Number of Training Instances Correctly Classified}}{\text{Number of Training Instances}}.$$

VII: Choose the attribute W found in Step III that has the second largest degree of entropy and calculate is highest and lowest attribute values based on the result obtained in Step I. Arrange them and subtract 0.05 cm fromthe attribute values that are at odd positions, and add 0.05 cm to the attribute values that are at even positions. Then, sort these values in an ascending sequence again.

VIII: Find the fuzzy rules whose classification error rate is larger than the level threshold value c, which is in the range [0, 1]. The classification error rate of a fuzzy rule is defined as follows:

$$\text{Classification Error Rate} = \frac{\text{Number of Training Instances Incorrectly Classified}}{\text{Number of Training Instances}}.$$

IX: Calculate the statistical distribution of the attribute values of the attribute W of the incorrectly classified training instances. If most of attribute values of these incorrectly classified training instances are falling in an interval belonging to the species Z2 and the corresponding linguistic term of this interval is Y2, then generate the following fuzzy rule:

IF X is Y 1 and W is Y 2 THEN the flower is Z2;

Modify the original generated fuzzy rule into the following form:

IF X is Y 1 and W is Y 2 THEN the flower is Z1;

where Y 2 denotes the complement of the fuzzy set Y2.

X: If there are fuzzy rules whose classification error rate is larger than the level threshold value c, then go to Step IX else go to Step XI.

XI: If the classification accuracy rate of the training instances is larger than the classification threshold value b, then Stop else go to Step VII.

5. NEURO FUZZY SYSTEM

In the field of artificial intelligence, **neuro-fuzzy** refers to combinations of artificial neural networks and fuzzy logic. Neuro-fuzzy was proposed by J. S. R. Jang. Neuro-fuzzy hybridization results in a hybrid intelligent system that synergizes these two techniques by combining the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks. Neuro-fuzzy hybridization is widely termed as Fuzzy Neural Network (FNN) or Neuro-Fuzzy System (NFS) in the literature. Neuro-fuzzy system incorporates the human-like reasoning style of fuzzy systems through the use of fuzzy sets and a linguistic model consisting of a set of IF-THEN fuzzy rules. The main strength of neuro-fuzzy systems is that they are universal approximators with the ability to solicit interpretable IF-THEN rules. The strength of neuro-fuzzy systems involves two contradictory requirements in fuzzy modeling: interpretability versus accuracy.

6. HEART DISEASE DATASET

In this section the details about the heart disease and its attributes taken for the experiment purpose is described. A database that contains the heart disease data is taken from UCI Machine learning repository. There are 303 samples or instances for 13 attributes like Thal {6—fixed defect,7—reversible defect,3—normal}, Number of major blood vessels colored by fluoroscopy {0,1, 2, 3}, Chest pain type {1, 2, 3,4}, Exercise induced angina {Yes, No}, Slope of peak exercise ST segment {Medium, High, Low} Oldpeak = ST depression induced by exercise relative to rest , Maximum heart rate achieved , Sex {Female Male},age, resting blood pressure, cholesterol. Out of the mentioned 13 attributes the relevant and best six attributes were found using the genetic algorithms. Genetic algorithm is an optimization technique based on the principles of genetics and natural selection. It uses a population of chromosomes to evolve under specified selection rules and operators like crossover and mutation. In feature selection, Genetic algorithms explore a large search space effectively. The obtained set of selected features include chest pain type, resting blood pressure, exercise induced angina maximum heart rate achieved, old peak and number of major vessels.

7. CONCLUSION

This paper is based on the heart disease diagnosis in patients with the help of evolutionary techniques like genetic algorithms, Fuzzy rule based expert system and Neural networks. This expert system will help the doctors to arrive at a conclusion about the presence or absence of heart disease in patients. This is an enhanced system that accurately classifies the presence of the heart disease. It is proved that the generated neuro-fuzzy rule based classification system is capable of diagnosing the heart disease in an effective and efficient method than any other classifiers.

REFERENCES

[1] Quinlan, J.R., (1986). Induction of decision trees, Machine Learning 1, 81–106.

[2] Quinlan, J.R., (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco.

[3] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., (1984). Classification and Regression Trees. Wadsworth, Belmont,CA.

[4] H. Liu and H. Motoda, (1998), Feature Selection for Knowledge Discovery and Data Mining, Boston: Kluwer Academic Publishers.

[5] M. Dash, H. Liu and H. Motoda, (2000), "Consistency Based Feature Selection," Proceedings of the Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining, pp. 98-109, Springer-erlag.

[6] S. Guha, R. Rastogi, and K. Shim, (1999). ROCK, "A Robust Clustering Algorithm for Categorical Attributes" Proceedings of the 15th International Conference on Data Engineering, Sydney, Australia,

[7] V. Ganti, J. Gehrke, and R. Ramakrishnan, (1999), "CACTUS-Clustering Categorical Data Using Summaries," Proceedings of the ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA.

[8] Tang, Z. H., MacLennan, J, (2005).: "Data Mining with SQL Server 2005", Indianapolis: Wiley.

[9] Kemal Polata, & Salih Güne¸sa & Sülayman Tosunb, (2007), Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing, ELSEVIER, PATTERN RECOGNATION.

[10] O.Cordon, F.Herrera, E.Herrera-ViedmaM.Lazano ,(1995) "Genetic algorithms and Fuzzy Logic in Control Processes" Technical Report #DECSAI-95109.

[11] Chang, C. H., & Chen, S. M. (2001). Constructing membership functions and generating weighted fuzzy rules from training data. In Proceedings of the 2001 ninth national conference on fuzzy theory and its applications, Chungli, Taoyuan, Taiwan, Republic of China (pp. 708–713).

[12] Chen, S. M., & Chang, C. H. (2005). A new method to construct membership functions and generate weighted fuzzy rules from training instances. Cybernetics and Systems, 36(4), 397–414.

[13] Chen, S. M., & Chen, Y. C. (2002). Automatically constructing membership functions and generating fuzzy rules using genetic algorithms. Cybernetics and Systems, 33(8), 841–862.

[14] Chen, S. M., & Fang, Y. D. (2005a). A new approach for handling the Iris data classification problem. International Journal of Applied Science and Engineering, 3(1), 37–49.

[15] Chen, S. M., & Fang, Y. D. (2005b). A new method to deal with fuzzy classification problems by tuning membership functions for fuzzy classification systems. Journal of Chinese Institute of Engineers, 28(1), 169–173.

[16] Chen, S. M., & Lin, H. L. (2005a). Generating weighted fuzzy rules for handling classification problems. International Journal of Electronic Business Management, 3(2), 116–128.

[17] Chen, S. M., & Lin, H. L. (2005b). Generating weighted fuzzy rules from training instances using genetic algorithms to handle the Iris data classification problem. Journal of Information Science and Engineering, 22(1), 175–188.

[18] hen, S. M., & Lin, S. Y. (2000). A new method for constructing fuzzy decision trees and generating fuzzy classification rules from training examples. Cybernetics and Systems, 31(7), 763–785.

[19] S. M., & Tsai, F. M. (2005). A new method to construct membership functions and generate fuzzy rules from training instances. International Journal of Information and Management Sciences, 16(2), 47–72.

[20] Y. C., Wang, L. H., & Chen, S. M. (2006). Generating weighted fuzzy rules from training data for dealing with the Iris data classification problem. International Journal of Applied Science and Engineering, 4(1), 41–52.

[21] R. (1936). The use of multiple measurements in taxonomic problem. Annals of Eugenics, 7, 179–188. Hong, T. P., & Chen, J. B. (1999). Finding relevant attributes and membership functions. Fuzzy Sets and Systems, 103(3), 389–404.

[22] T. P., & Lee, C. Y. (1996). Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems, 84(1), 33–47. Hong, T. P., & Lee, C. Y. (1999). Effect of merging order on performance of fuzzy induction. Intelligent Data Analysis, 3(2), 39–151.

[23] F. M., & Chen, S. M. (2002). A new method for constructing membership functions and generating fuzzy rules for fuzzy classification systems. In Proceedings of 2002 tenth national conference on fuzzy theory and its application, Hsinchu, Taiwan, Republic of China.

[24] P., & Chen, S. M. (1999). A new method for constructing membership functions and fuzzy rules from training examples. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 291), 25–40.

[25] L. A. (1965). Fuzzy sets. Information and Control, 8, 338–353.Zadeh, L.A., "Fuzzy sets," Information and Control, Vol. 8, pp. 338-353, 1965.