A Fuzzy Rule Based Expert System for Effective Heart Disease Diagnosis

E.P. Ephzibah[1] and 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. This is a general method that combines the soft computing techniques like genetic algorithms and fuzzy rule based expert system for effective heart disease diagnosis. It is very important to diagnose the disease in the early stage itself. Prompt and correct diagnosis of the disease by selecting the important and relevant features will help to discard irrelevant and unimportant ones. Genetic algorithms help in feature subset selection. After the subset selection the fuzzy rule based expert system provides the classificatory knowledge. The proposed system generates the rules from the instances and narrows down the limit of the rules using degree of the memberships. The system is designed in Matlab software. The system can be viewed as an alternative method for effective diagnosis of heart disease presence.

Keywords: Fuzzy rule based Expert system, genetic algorithm, heart disease diagnosis.

1 Introduction

One of the leading causes of death in the world is heart disease. Especially in India many people die due to the ignorance of the disease and also due to lack of proper diagnostic methods. It will be of greater value if the disease is diagnosed in its early stage. Prompt and correct diagnosis of the disease will definitely decrease the death rate due to heart failures. There are many tests that are available for the diagnosis of this disease, but in particular they may take more time and also cost more. This paper is proposed to find out the required and relevant tests that should be done on the patients' body to make the diagnosis process faster as well as less expensive. Feature subset selection is a methodology which reduces the features into a subset and subset alone can be considered for further study in the diagnosis process. Classification plays a very important role after selecting the feature subset. The required data is given to the classifier for further classification by dividing the

N. Meghanathan et al. (Eds.): CCSIT 2012, Part II, LNICST 85, pp. 196–203, 2012.

samples into training and testing samples. Finally the accuracy is checked, which is a measuring agent for the efficiency of the selected features. In this paper a fuzzy expert system is framed with efficient set of rules to determine the presence or absence of heart disease.

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.

The paper is organized in the manner that in section 3 the theory on soft computing has been explained, in section 4 a brief description on genetic algorithms is given , sections 5,6,7 deal with the fuzzy expert system and the sequential steps involved in fuzzy rule generation, section 8 illustrates the heart disease dataset used for our work, section 9 clearly explains the rules that are generated for our work and finally section 10 gives the conclusion.

3 Soft Computing

Soft computing is a methodology that tends to combine the different aspects of fuzzy logic, evolutionary algorithms like genetic algorithms neural networks, and non-linear distributed systems. It is a way to implement hybrid systems that helps to get innovative solutions in the various sectors of intelligent control, modeling and classification.

Soft Computing are Fuzzy Logic, Neural Computing, Evolutionary Computation, Machine Learning and Probabilistic Reasoning and parts of learning theory. Soft computing is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in soft computing are complementary rather than competitive. Soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence.

4 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 pay 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.

5 Fuzzy Expert System

The general fuzzy inference process proceeds with the following four steps.

1. **FUZZIFICATION:** Fuzzification is a process of finding the membership value of a scalar in a fuzzy set. This is done by fuzzifying the input values. To obtain the degree of truth for the rules generated, the membership function is applied on the input values.

2. **FUZZY INFERENCE:** A fuzzy inference engine consists of assessment, correlation, aggregation and reduction. Assessment is the process of gathering initial evidence and determining the truth of the rule antecedent. The amount of evidence of the rule is assessed by combining the individual fuzzy propositions. Correlation matches the truth of the consequent. It takes the form of either scaling or truncating the consequent using the truth. Aggregation updates the current outcome fuzzy set using the correlated consequent fuzzy set. Each rule updates the outcome fuzzy set with its own correlated fuzzy sets. Two methods of aggregation are addition and maximization. This process plays a very important role as heart of the fuzzy inference engine.

3. **DEFUZZIFICATION:** It is a process of reduction that produces a scalar result from the final outcome fuzzy set. There are many methods of defuzzification, but the most commonly used is centroid or center of gravity. This is calculated as the weighted average of the outcome fuzzy set.

6 Rule-Based Fuzzy Expert Systems

One of the most important components in the development of expert system is knowledge acquisition. This could be obtained by questionnaires, interviews and experiences. To find a solution computer programs use well- structured algorithms and some reasoning strategies. For expert systems it may be more useful to employ heuristics: strategies that often lead to the correct solution. Human intelligence is applied to solve conventional rule-based expert systems. Rule based expert systems play a very important role in modern intelligent systems. A rule is a statement that combines or links the input scalar to an output vector.

In the above figure the knowledge base contains all the information, relationships and related data. The inference engine obtains the knowledge from the knowledge base and uses that knowledge for predictions and results. The explanation facility helps in explaining the user how the results are obtained. The purpose of knowledge acquisition facility is to provide the way for capturing, storing all the components of knowledge base. A rule-based system consists of if-then rules with facts. These if-then rules are use to generate conditional statements.

7 Heart Disease Data-Set

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.

8 Fuzzy Rule Generation

In a fuzzy rule-based system the rules can be represented in the following way:

If (x is A) and (y is B) … and…. Then (z is Z)

where **x,y** and **z** represent variables (or attributes or features) and **A, B** and **Z** are linguistic variables such as `far', `near', `small'. Ishibuchi *et al* [11] [12] [10] report various results when using GAs to select fuzzy rules in classification. The problem tackled is that of using a set of numerical data to generate if-then rules for pattern classification. There are some steps to be followed for the fuzzy rules to be generated. From the heart disease data set we take a set of input and output attributes and classes. There are around thirteen attributes and we have reduced them into six relevant input attributes and one output attribute, y. Even though there are four classes in the output as 0,1,2,3,4, this work takes them as only two classes as {0,1}. The class 0 represents that the patient has no heart disease and class 1 represents the presence of the heart disease. We name them as follows:

$$\{ x_1, x_2, x_3, x_4, x_5, \quad x_6\},\{y\}$$

The task of generating the fuzzy rules involves few steps that follow [2]:

Step1: Separate the input space into fuzzy regions:

Find the maximum attribute value and the minimum attribute value of each attribute of the training instances. This shows that the value of the particular attribute is in this interval. Divide the domain interval of each attribute into 2N+1 region. The following table gives the range of the attributes:

S.No	Attributes	Maximum	Minimum
1	chest pain type	4	1
2	resting blood pressure	200	94
3	maximum heart rate achieved	202	71
4	exercise induced angina	1	0
5	Oldpeak	6.2	0
6	number of major vessels	3	0

Fig. 1. This show the maximum attribute vale and minimum attribute vales

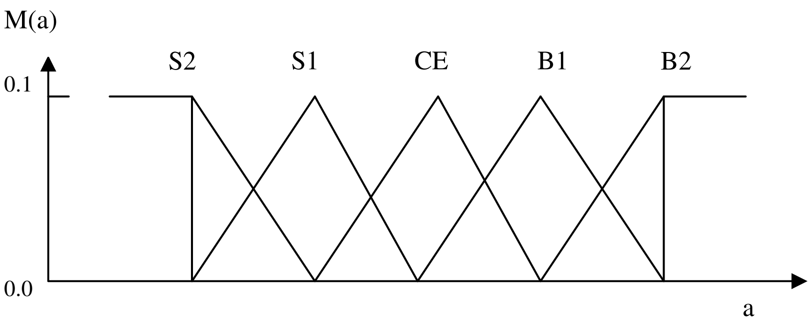

Fig. 2. Division of input spaces into fuzzy regions

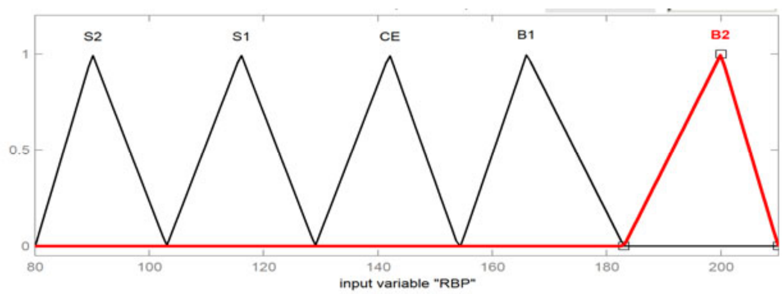

Fig. 3. Fuzzy regions for resting blood pressure attribute

Step2: Develop fuzzy rules from the example set:

The fuzzy rules are generated based on the attribute values of each and every instance. The selected attributes are named as x_1, x_2, x_3, x_4, x_5, and x_6 for chest pain type, resting blood pressure, maximum heart rate, exercise induced angina, old peak and number of major blood vessels colored by fluoroscopy value respectively. These are some of the rules obtained form the example instances.

Instance 1:
If (x_1 is 1) and (x_2 is high) and (x_3 is medium) and (x_4 is 0) and (x_5,is risk) and (x_6 is 0) then result is 0.
Instance 2:
If (x_1 is 4) and (x_2 is very high) and (x_3 is medium) and (x_4 is 1) and (x_5 is risk) and (x_6 is 3) then result is 1.
Instance 3:
If (x_1 is 4) and (x_2 is low) and (x_3 is medium) and (x_4 is 1) and (x_5 is terrible) and (x_6 is 2) then result is 1.
Instance 4:
If (x_1 is 3) and (x_2 is medium) and (x_3 is medium) and (x_4 is 0) and (x_5 is terrible) and (x_6 is 0) then result is 0.

Instance 5:
If (x_1 is 2) and (x_2 is medium) and (x_3 is medium) and (x_4 is 0) and (x_5 is risk) and (x_6 is 0) then result is 0.

There rules are generated based on the membership values and their region. Like this the rules are generated. Based on the number of instances the number of rules may change.

Step3: Select the best rule based on the degree of the rule:

As there are many rules generated there are more chances of conflicts and redundancy. For example, many rules can have the same IF part but different THEN part. This is where the conflicts occur and to avoid the conflicts the rules are ranked based on their degrees. The degree of a rule is obtained based on the membership values. Sum of all the membership values is obtained and accordingly the rules are selected. If two rules have the same IF part, the rule with higher degree of membership is selected and the other one is discarded. Thus a limited set of definite rules are generated. Finally mapping the input and output values using the generated fuzzy rule base is done to check for any classification.

9 Conclusion

In this paper, we have given a detailed presentation of fuzzy rule-based expert system for effective diagnosis of the heart disease in patients. This expert system will help the doctors to arrive at a conclusion about the heart disease in patients. Our expert system is an enhanced system that accurately classifies the presence of the heart disease. It is shown that the generated fuzzy rule based classification system is capable of diagnosing the heart disease than any other classifiers.

References

1. Chang, C.H., Chen, S.M.: 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 (2001)
2. Chen, S.M., Chang, C.H.: A new method to construct membership functions and generate weighted fuzzy rules from training instances. Cybernetics and Systems 36(4), 397–414 (2005)
3. Chen, S.M., Chen, Y.C.: Automatically constructing membership functions and generating fuzzy rules using genetic algorithms. Cybernetics and Systems 33(8), 841–862 (2002)
4. Chen, S.M., Fang, Y.D.: A new approach for handling the Iris data classification problem. International Journal of Applied Science and Engineering 3(1), 37–49 (2005a)
5. Chen, S.M., Fang, Y.D.: 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 (2005b)

6. Chen, S.M., Lin, H.L.: Generating weighted fuzzy rules for handling classification problems. International Journal of Electronic Business Management 3(2), 116–128 (2005a)
7. Chen, S.M., Lin, H.L.: 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 (2005b)
8. Chen, S.M., Lin, S.Y.: A new method for constructing fuzzy decision trees and generating fuzzy classification rules from training examples. Cybernetics and Systems 31(7), 763–785 (2000)
9. Chen, S.M., Tsai, F.M.: 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 (2005)
10. Chen, Y.C., Wang, L.H., Chen, S.M.: 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 (2006)
11. Fisher, R.: The use of multiple measurements in taxonomic problem. Annals of Eugenics 7, 179–188 (1936); Hong, T.P., Chen, J.B.: Finding relevant attributes and membership functions. Fuzzy Sets and Systems 103(3), 389–404 (1999)
12. Hong, T.P., Lee, C.Y.: Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems 84(1), 33–47 (1996); Hong, T.P., Lee, C.Y.: Effect of merging order on performance of fuzzy induction. Intelligent Data Analysis 3(2), 39–151 (1999)
13. Ishibuchi, H., Nakashima, T.: Effect of rule weights in fuzzy rulebased classification systems. IEEE Transactions on Fuzzy Systems 9(4), 506–515 (2001)
14. Tsai, F.M., Chen, S.M.: 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 (2002)
15. Wu, T.P., Chen, S.M.: A new method for constructing membership functions and fuzzy rules from training examples. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 29(1), 25–40 (1999)
16. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
17. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, and 3. Information Sciences, 8:199–249, 8:301–357, 9:43–80 (1975)