Framing Fuzzy Rules using Support Sets for Effective 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

Significance and relevance of certain features are obtained by various techniques. Feature subset selection involves summarizing mutual associations between class decisions and attribute values in a pre-classified database. In this paper genetic algorithm is used to find the relevant set of features by optimizing the fitness function and using the operators like crossover and mutation. Fuzzy logic is a form of knowledge representation suitable for notions that cannot be defined precisely, but which depend upon their contexts. In this work the fuzzy rules are framed with the help of support sets. The classification done using fuzzy inference system provides results that are better than other techniques.

Keywords

Genetic Algorithms, Fuzzy logic, Medical data, Disease diagnosis.

1 INTRODUCTION

Genetic algorithms are known to be one of the best methods for search and optimization problems. Feature selection is the task of identifying and selecting a useful subset of pattern-representing features from a large set of features. The number of features increases as the dimensionality expands. The benefits of feature selection are facilitating data visualization and understanding, reducing the measurement and storage requirements, reducing training time, defying the curse of dimensionality to improve prediction performance. The objective is finally to construct and select subsets of features that are useful to build a good predictor [1]. A subset of useful features may exclude redundant but relevant features. After generating the best feature subset it is used for classification. Fuzzy set theory and fuzzy logic are highly suitable for developing knowledge based systems in medicine for diagnosis of diseases. In this paper it is discussed about how genetic algorithms and fuzzy logic combine together for efficient and cost effective diagnosis of heart disease.

DOI : 10.5121/ijfls.2012.2102

This paper is organized as follows. In section 2 related works is explained. In section 3 the proposed technique is described. In section 4 framing fuzzy rules have been explained and in section 5, 6 and 7 experimental results with system and conclusion are given.

2 RELATED WORK

Experimental results have shown that genetic algorithms are able to reach a relative good score in a quite small number of generations, for function optimization. They refine the solution space trying to identify the exact optimal solution of the function. There are a good number of methods which reached high classification accuracies using the dataset taken from UCI machine learning repository. Among these, [2] Tool Diag, RA obtained 50.00% classification accuracy by using IB1- 4 algorithm. [2] WEKA, RA obtained a classification accuracy of 58.50% using InductH algorithm while ToolDiag, RA reached to 60.00% with RBF algorithm. [2] Again, WEKA, RA applied FOIL algorithm to the problem and obtained a classification accuracy of 64.00%. [2] MLP+BP algorithm that was used by ToolDiag, RA reached to 65.60%. [2] The classification accuracies obtained with T2, 1R, IB1c and K* which were applied by WEKA, RA are 68.10%, 71.40%, 74.00% and 76.70%, respectively. [2] Robert Detrano used logistic regression algorithm and obtained 77.0% classification accuracy. [15] According to Sellappan Palaniappan, Rafiah Awang, Naïve Bayes gives the highest probability (95%) with 432 supporting cases, followed by Decision tree (94.93%) with 106 supporting cases, and Neural Network (93.54%) with 298 supporting cases. Naïve Bayes appears to be most effective as it has the highest percentage of correct predictions (86.53%) for patients with heart disease, followed by Neural Network (with a difference of less than 1%) and Decision Trees.

3 PROPOSED METHOD

The proposed method is an extended version of the model that combines the genetic algorithms for feature selection and fuzzy expert system for effective classification. In this paper the rules are generated based on the support sets found. The dataset from UCI machine learning repository is used to diagnose the presence of heart disease based on the various medical tests carried out on a patient. It contains elements of two classes: patients with and without heart disease. There are about 303 instances and 76 attributes. Only 14 attributes out of 76 has been identified to be effective and necessary attributes. According to the proposed method only six attributes are considered. Genetic algorithm is a class of optimization procedure used to solve problem that involves a search to find the optimal solution. GAs operate iteratively on a population of chromosomes, each one of which represents a candidate solution to the problem at hand, properly encoded as a string of symbols(e.g. binary). A randomly generated set of such chromosomes form the initial population from which the GA starts its search. Three basic genetic operators guide this search: selection, crossover, and mutation. The genetic search process is iterative: evaluating, selecting, and recombining chromosome string in the population during each iteration (generation) until reaching some termination condition. Evaluation of each chromosome is based on a fitness function. It determines which of the candidate solutions are better. The GA combines selection, crossover, and mutation operators with the goal of finding the best solution to the problem by searching until the specified criterion is met.

A fuzzy set is a collection of distinct elements with a varying degree of relevance or membership. The membership function takes interval values between 0 and 1. These values express the degrees with which each object is compatible with the properties or features that are distinctive to the collection. A fuzzy set is a generalization of the concept of a set whose characteristic function takes only binary values. A fuzzy inference model can be created using the properties of fuzzy set. The knowledge base of a fuzzy inference system is to link the fuzzified inputs with the associated reasoning mechanism.

There are two major models of fuzzy system, Mamdani [4] and Takagi-Sugeno (T-S) [5] fuzzy systems. The main difference between these two types of fuzzy systems lies in the consequent variable of fuzzy rules. Mamdani type fuzzy systems use linguistic fuzzy sets as consequent variables in fuzzy rules, whereas the T-S type fuzzy systems employ a linear combination of input variables as a rule consequent variable. This work has been implemented using Mamdani type.

Based on the experts' (Doctors') knowledge the fuzzy rules were generated. The generated rules help us to predict the disease using fuzzy tool in Matlab. The input is the set of all the selected features and the output of the system is to get a value 1 or 0 that indicates the presence or absence of the disease.

4 FRAMING FUZZY RULES USING SUPPORT SETS

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. In Fuzzy Logic, the truth of any statement is a matter of degree. In order to define Fuzzy Logic operators, we have to find the corresponding operators that preserve the results of using *AND, OR,* and *NOT* operators. The answer is *min, max,* and *complements* operations. These operators are defined, respectively, as

$$\mu_{A\cup B}(x) = \max[\mu_A(x), \mu_B(x)] \text{ -------------(1)}$$
$$\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)] \text{ -------------(2)}$$
$$\mu_A(x) = 1 - \mu_A(x) \text{ -------------(3)}$$

Zadeh in 1965 defined fuzzy union and fuzzy intersection as

$$\mu_{A\cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\,\mu_B(x) \text{ -------------(4)}$$
$$\mu_{A\cap B}(x) = \mu_A(x)\,\mu_B(x) \text{ -------------(5)}$$

Fuzzy inference systems consist of if–then rules that specify a relationship between the input and output fuzzy sets. Fuzzy relations present a degree of presence or absence of association or interaction between the elements of two or more sets. Let *U* and *V* be two universes of discourse. A fuzzy set is a relation R(U,V) in the product space U x V and is characterized by the membership function $\mu_R(x,y)$, where $x \in U$ and $y \in V$ and $\mu_R(x,y) \in [0,1]$. Fuzzy relations play an important role in fuzzy inference systems. FL uses notions from crisp logic. Concepts in crisp logic can be extended to FL by replacing 0 or 1 values with fuzzy membership values. A singleton fuzzy rule assumes the form "if *x* is *A*, then *y* is *B*," where $x \in U$ and $y \in V$, and has a membership function $\mu_{A \rightarrow B}(x,y)$, where $\mu_{A \rightarrow B}(x,y) \in [0,1]$. The if part of the rule, "*x* is *A*," is called the *antecedent* or *premise*, while the then part of the rule, "*y* is *B*," is called the *consequent* or *conclusion*. Interpreting an if–then rule involves two distinct steps. The first step is to evaluate the antecedent, which involves fuzzifying the input and applying any necessary fuzzy operators. The second step is implication, or applying the result of the antecedent to the consequent, which essentially evaluates the membership function $\mu_{A \rightarrow B}(x,y)$. It can be seen that in crisp logic a rule is fired if the premise is exactly the same as the antecedent of the rule, and the result of such rule firing is the rule's actual consequent. In fuzzy logic, a rule is fired so long as there is a nonzero degree of similarity between the premise and the antecedent of the rule. For most applications, the fuzzy membership function $\mu_{A \rightarrow B}(x,y)$ for a given relation is obtained with the minimum or product implication, given, respectively, as follows:

$$\mu_{A\cap B}(x) = \mu_A(x), \mu_B(x) \qquad \text{----------------(6)}$$
$$\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)] \qquad \text{----------------(7)}$$

It was Mamdani (1977) who first proposed the minimum implication, and later Larsen (1980) proposed the product implication. The minimum and product inferences have nothing to do with traditional prepositional logic; hence, they are collectively referred to as engineering implications. Details of implication methods can be found in the classic tutorial paper by Mendel (1995).

A fuzzy rule base contains a set of fuzzy rules *R*. A single if–then rule assumes the form "if *x* is *Tx* then *y* is *Ty* ." An example of a rule might be "if *education* is *high* and *experience* is *high*, then *salary* is *very high*." For a multiinput and multioutput system, the rules are represented as

$$R=(R_1,R_2,R_3,\ldots R_n).$$

This proposed system generates the fuzzy rules based on the support sets obtained.

SNo	Attributes	Support set	
		Heart patients	Non-heart patients
1	Chest Pain Type	4	1,2,3
5	Rbps	134-153	142-154
2	Exang	yes	No
3	Oldpeak	2.06-6.2	<2.06
4	Thalach	71-136	136-168
6	ca	1,2,3	0

Table1. Table shows the values of the features in the support set.

5 EXPERIMENTAL RESULTS

The process of fuzzy logic is that: Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step. In this work the list of attributes selected by genetic algorithm using MATLAB are given as input to the fuzzy system and the fuzzy linguistic terms and membership functions are defined based on the existing knowledge and the experts advice. The output field refers to the presence of heart disease in the patient.

About the features and their fuzzy values:

The first feature considered is the chest pain type for which there are 4 values like typical angina, atypical angina, non-anginal pain and asymptomatic. The asymptotic pain is the support set for heart patients where as the first three pain types are in the support set for non-heart disease patients. The next feature is the resting blood pressure. There are two fuzzy regions namely, "low" and "high". The value for high region belongs to the support set for heart patients. The next feature in the list is exercise induced angina, for which there are two values like "yes" or "no". The value "yes" is in the support set for patients with heart diseases. Old peak has the fuzzy region like low and high and the support set for heart patients is high. Maximum Heart rate is the next feature in the list that has 2 linguistic variables or fuzzy sets namely, Low and High. The support set is low for the patients with heart disease and high for non-heart disease patients. The last feature in the subset is the no. of vessels colored. For this feature there are four fuzzy sets zero, one, two and three. The support set show values 1,2,3 for heart patients and 0 for non-heart patients.

6 SYSTEM TESTING

The output shows the presence or absence of heart disease in a patient given the values for the input features. There are two fuzzy sets, "healthy" & "sick". The membership functions of "healthy" and "sick" are trapezoidal.

The picture given below is the rule viewer: heart disease. Given an instance for which the result is "sick", the rule viewer correctly points out the presence of the disease.

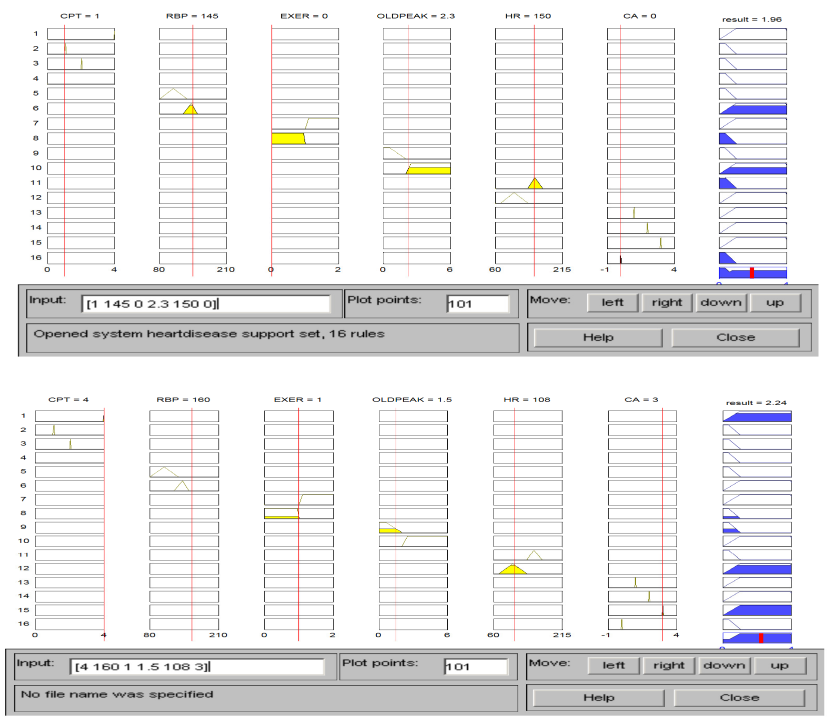

Fig1&2: Figures show the rule viewer that represents the input and output values.

Chest Pain Type	Resting blood pressure	Exercise induced angina	Oldpeak:	Maximum heart rate achieved	No. of major vessels	Predicted Output
1	145	0	2.3	150	0	"Healthy"
4	160	1	1.5	108	3	"Sick"

Table 2: System Testing

7 CONCLUSION

In this paper a system that combines genetic algorithms and fuzzy expert system is proposed. Genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient. Designing of this system with fuzzy in comparison with other methods improves results. The experts knowledge and well as the support sets have been used in framing the fuzzy rules. The explained model proves to be more efficient in diagnosing heart disease. Furthermore, it has been proven to be competitive with state of the art classifiers like Naïve Bayes, Decision tree, Classification via clustering and SVM classifier.

REFERENCES

1. Isabelle Guyon, Andre Elisseeff: An introduction to variable and feature selection, Journal of Machine learning research 3 1157-1182, (2003)
2. Kemal Polata, Salih Gune¸sa & Sulayman Tosunb: Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing, Elsevier , Pattern recognation,(2007)
3. Robert Detrano, M.D, PhD.: V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.
4. Zimmermann, H.-J.: "Fuzzy Set Theory - And its Applications," 3rd Ed, Kluwer Academic Publishers, (1997)
5. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control, IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, pp.116-132, Jan. (1985)
6. Zadeh, L.A.: Fuzzy Sets. Information and Control, Vol.**8,** pp.338-353 (1965).
7. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley. (1989)
8. Novruz Allahverdi, Serhat Torun & Ismail Saritas: Design of a fuzzy expert system for determination of coronary heart disease risk, International Conference on Computer Systems and Technologies - CompSysTech'07.
9. Kwong, C. K., Chang, K.Y. & Tsim, Y. C.: A genetic algorithm based knowledge discovery system for the design of fluid dispensing processes for electronic packaging, Expert Systems with Applications, 36(2), 3829–3838. (2008)
10. J.Shapiro: Genetic algorithms in machine learning. Lecture notes in computer science (Vol. 2049/2001, pp. 146–168). (2001)
11. Zhu, F., & Guan, S.: Feature selection for modular GA-based classification. Applied Soft Computing, pp.381–393.(2004)
12. Booker, L. B., Goldberg, D. E., & Holland, J. H.: Classifier systems and genetic algorithms, Artificial Intelligence, 40(1–3), 235–282.(1989)
13. Soler, V., Roig, J., Prim, M.: Finding Exceptions to Rules in Fuzzy Rule Extraction, KES 2002, Knowledge-based Intel. Information Engineering Systems, Part 2,1115-1119.(2002)
14. Latha Parthiban, Subramanian, R:Intelligent heart disease prediction system using CANFIS and Genetic Algorithm, International Jouranl of Biological Life Sciences, (2007)
15. Sellappan Palaniappan, Rafiah Awang "*Intelligent Heart Disease Prediction System Using Data Mining Techniques*" International Conference on Computer Systems and Applications, April 2008. pp.108-115, AICCSA 2008. IEEE/ACS.(2008)