Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 172-177

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

An Expert System for Breast Cancer Diagnosis Using Fuzzy Classifier with Ant Colony Optimization

[1]Anto S and [2]Chandramathi S

[1]Department of CSE, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.
[2]Department of ECE, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.

ARTICLE INFO

Article history:
Received 20 January 2015
Accepted 02 April 2015
Published 20 May 2015

Keywords:
Breast Cancer Diagnosis
Ant Colony Optimization
Expert System Fuzzy Inference

ABSTRACT

Medical diagnosis is a complex process in which the result of the diagnosis has to be more accurate. Medical expert systems can assist physicians to make fast, accurate and meaningful clinical decisions. This paper proposes a medical decision support system based on fuzzy logic and Ant Colony Optimization (ACO) for the breast cancer diagnosis using Wisconsin's breast cancer dataset of UCI machine learning repository. A set of fuzzy rules are extracted from patient dataset using Fuzzy Logic and ACO. The ACO algorithm optimizes these extracted fuzzy rules and generates optimized set of rules. The fuzzy inference system uses these optimized rules to perform classification of the test data. Ten-fold cross validation procedure is used to evaluate the performance of the system in terms of the classification accuracy. The results show that the proposed model achieves better accuracy with other existing systems in the literature.

To Cite This Article: Anto S and Chandramathi S., An Expert System for Breast Cancer Diagnosis Using Fuzzy Classifier with Ant Colony Optimization. ***Aust. J. Basic & Appl. Sci.,*** *9(16): 172-177, 2015*

INTRODUCTION

Machine learning is a division of Artificial Intelligence (AI) which facilitates the computing systems to build the ability to learn from data and to make autonomous decisions. It is generally categorized into two types namely supervised learning or unsupervised learning. Supervised learning is the machine learning technique of inferring a pattern from a set of training data that have labels. On the other hand, unsupervised machine learning algorithms use unlabeled data instances to extract interesting pattern from the dataset. Medical expert system for diagnosis of different diseases has received more attention for past few years. Knowledge discovery in patient's data and machine learning are used to design such expert systems. These decision support systems can play major role in assisting the physicians while making complex clinical decisions thereby can improve the accuracy of diagnosis.

During recent years there is a sizeable increase in the number of cases on breast cancer all over the world. It is reported that breast cancer is the second most diagnosed cancer type. Early detection and flawless diagnosis of this cancer disease can help the doctors to treat it properly and hence can reduce the death count. With the development of more effective diagnostic methods and tools like medical expert systems and improvements in treatment methodologies have led to the save of considerable number of lives. The long-term survival duration for patients in whom breast cancer has not dangerously transformed has increased, with the majority of women surviving more number of years after proper diagnosis and treatment (Polat and Gunes, 2007). The proposed work uses an ACO based classifier to extract a set of fuzzy rules from the dataset for the diagnosis of breast cancer. The final result of the system is produced based on the classification accuracy according to training and testing.

The Fuzzy set theory presents a general method to obtain a form of linguistic IF-THEN rules. The IF-THEN rules that are generated from the numerical medical data or expert experiences (knowledge) determine the performance of the fuzzy system. Generally, a fuzzy classification-inference system has two main characteristics- (1) the number of fuzzy partitions (2) the number of IF-THEN rules. So the usage of more number of rules may increase the classification accuracy of the system but can lead to the increase of the computational complexity. Fuzzy approaches in the classification task make fast and accurate decisions. ACO-based algorithms were developed by the inspiration of the natural behavior of ants to follow the shortest path while hunting for food based on the density of pheromone trails. Upon finding food the ants return to their starting place by

Corresponding Author: Anto S, Department of CSE, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.

leaving down pheromones which will be followed by other ants based on pheromone density. Thus they follow the shortest – optimized path from the colony to food source (Martens *et al*., 2007). So this nature of ant colonies can be used to attack optimization problems by developing stochastic search heuristics based on probability density functions. In this proposed work, the ACO algorithm is used to generate rules from the medical dataset and to select optimal number of fuzzy IF-THEN rules from them. These selected rules are used by fuzzy inference system to perform classification.

2 Related Works:

(Huang *et al*., 2006) have proposed a Genetic Algorithm (GA) based methodology for the feature selection and parameters optimization of Support Vector Machines (SVMs). It focuses on optimizing the kernel parameters and feature subset selection without degrading the SVM classification accuracy. 94.23% accuracy is obtained.

(Polat *et al*., 2007) have used Least Square Support Vector Machine (LS-SVM) for Breast cancer diagnosis. In contrast to the previously reported classification techniques, the obtained accuracy is very promising.

(Nauck and Kruse ,1999) have designed a classifier to generate interpretable fuzzy classification rules from medical data. Extensions are applied to the neuro-fuzzy learning algorithms of this classification system. The accuracy of the system is 95.06%.

(Ubeyli, 2007) have propounded diagnostic systems for detecting breast cancer, provides an integrated view of implementing automated cancer diagnostic system. 91.92% accuracy is obtained, which is little lower in contrast to the existing diabetes diagnosis systems.

(Karabatak and Ince ,2009) have developed an Association Rule (AR) and Neural Network (NN) based breast cancer detection expert system. In this work , the dimensions of breast cancer dataset are reduced using AR and intelligent classification is performed using NN. The classification rate is reported to be 95.6%.

The proposed system uses ACO algorithm to generate fuzzy classification rules from training patters of the dataset. The artificial ants make candidate fuzzy rules gradually in search space. The stochastic behavior of ACO algorithm encourages the ants to find more accurate rules. These optimized rules are used by the fuzzy inference engine to perform decision making on testing patters as shown in Figure 1.

3 Proposed Scheme:

ACO system uses artificial ants that work with a probability density function and produce solutions to optimization problems (Dorigo and Blum, 2005). These software agents mimic the food hunting behavior of their natural counterparts in finding the shortest-path to the food source from the colony.

The Ant System (Dorigo *et al*.,1991), (Dorigo *et al*.,1996) is the first algorithm that uses the principles of the ACO heuristic. The ants iteratively construct solutions and add pheromone to the paths corresponding to these solutions. The stochastic procedure of path selection is based on two parameters, the heuristic values and pheromone. The quantity of ants that chose the trail recently is given by pheromone value, and the value of the heuristic is a problem dependent quality measure. The ants choose the trail with the highest pheromone density and heuristic behavior.

Once the ant arrives at its destination, the solution of the path followed by the ant is evaluated and the pheromone value is expanded appropriately (Ganji and Abadeh, 2011). Evaporation gradually reduces the pheromone level of all trails. Consequently, the pheromone levels of the trails that are not followed gradually decrease, which in turn lower the probability of the trail being chosen by subsequent ants.

3.1 Membership value assignment by normalization of dataset:

In this work, the Wisconsin (original) breast cancer dataset is used. The dataset has a total of 699 records and each record has 9 attributes. The attributes are graded from ‘1’ (Normal State) to ‘10’ (Most Irregular State).The grading of the attributes is given in Table 1. In this database, 241 (65.5%) records are found to be malignant and 458 (34.5%) records are benign.

Normalization of dataset between 0.0 and 1.0 is done using the min-max normalization method (equation 1)

$$\text{Normalize}(X) = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where, (X)=membership function
X= linguistic value
X_{min}= least linguistic value
X_{max}=maximum linguistic value

The domain of each attribute is homogeneously partitioned into symmetric triangular fuzzy sets. The Membership function of each linguistic value is determined from the domain as shown in figure 2. The used antecedent fuzzy are S: Small, MS: medium small, M: medium, ML: medium large, L: large.

Australian Journal of Basic and Applied Sciences, 9(13) Special 2015, Pages: 172-177

Table 1: Description of attributes of the dataset

Attribute No.	Values	Standard Deviation	Mean
1	1-10	2.82	4.44
2	1-10	3.07	3.15
3	1-10	2.99	3.22
4	1-10	2.86	2.83
5	1-10	2.22	3.23
6	1-10	3.64	3.54
7	1-10	2.45	3.45
8	1-10	3.05	2.87
9	1-10	1.73	1.60

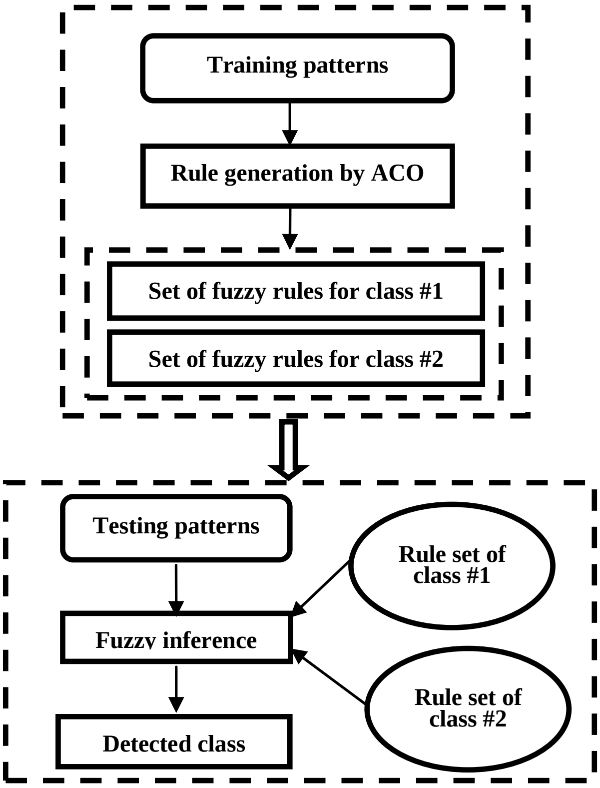

Fig. 1: Stages of the Proposed System

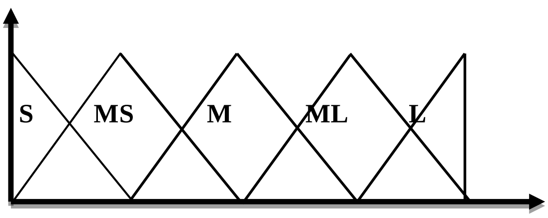

Fig. 2: Antecedent fuzzy sets

3.2 Rule generation:

In training stage a set of fuzzy rules are generated using the training patterns. The fuzzy rules are displayed as the following form:

Rule R_j If x_1 is A_{j1} and … and x_n is A_{jn}, then Class C_j with $CF=CF_j$.

where R_j is the label of jth fuzzy if-then rule, $A_{j1}\ldots A_{jn}$ are antecedent fuzzy sets on the unit interval [0,1], C_j is the consequent class, CF_j and is the grade of certainty of the fuzzy If-then rule.

The rule learning process is done separately for each class. The list of discovered rule is initially empty and the training samples consist of all the training samples. The first ant constructs the rule Rj randomly by adding one term at a time. The ants modify the rule Rj according to the Max_change parameter.

3.2.1 Rule Modification:

During initial iteration (t=0), ant_0 creates rule and in subsequent iteration ($t \geq 1$) the ants will modify the rule. The maximum number of times that each ant can modify the rule in each iteration ($t \geq 1$) is decided by a parameter named Max_change. Each ant chooses $term_{ij}$ to modify according to following probability,

$$P_{ij} = \frac{\tau_{ij}(t).\eta_{ij}}{\sum_i^a \sum_j^{bi} \tau_{ij}(t).\eta_{ij}}, \forall i \in I \quad (2)$$

where $\tau_{ij}(t)$= pheromone currently available on the path between attribute i and antecedent fuzzy set j.

a = total number of attributes

b_i=total number of antecedent fuzzy set for $attribute_i$

l =the set of attributes that are not yet used by the ant.

Corresponding to the quality of the modified rules, pheromone is assigned to each trail path of the ant. The ants choose the path in which the pheromone trail is high.

3.2.2 Heuristic information:

The ants modify the rule using the Heuristic information and the amount of pheromone. ANTMINER+ (Martens *et al*., 2007) uses a set of two dimensional matrices as heuristic information for each class. The rows represent the attributes and the columns representing the fuzzy values. These matrices help the ants to choose more accurate rules.

3.2.3 Pheromone update rule:

Pheromone updating is carried out only when the there is a improvement in the quality of the modified rule when compared to the quality of the rule before modification. Pheromone updating is carried out using the following equation.

$$\Delta_Q = Q_i^{\text{AFTER MODIFICATION}} - Q_i^{\text{BEFORE MODIFICATION}} \quad (3)$$

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t)(\Delta_i^Q.C) \quad (4)$$

where Δ_i^Q shows difference between the quality of rule after and before modification and C is parameter to regulate the influence of Δ_i^Q to increase the pheromone.

3.3 Fuzzy inference:

The ACO- Fuzzy system has generated and optimized a set of fuzzy rules in order to classify the test data, first certainty grade of each fuzzy rule is calculated. To compute the certainty grade of each fuzzy If-then rule the following steps are followed.

Step 1: The compatibility of each training pattern $x_p=(x_{p1},x_{p2},\ldots x_{pn})$ with the fuzzy If-then rule R_j by the following product operation.

$$\mu_j(X_p) = \mu_{j1}(X_{p1}) \times \ldots \mu_{jn}(X_{pm}),\ p = 1, 2, 3, m \quad (5)$$

where $\mu_{ji}(x_{pi})$ = Membership function of ith attribute of pth pattern.

m = Total number of patterns.

Step 2: Relative sum of compatibility grades of the training pattern is computed with each fuzzy If-then rule R_j.

$$\beta_{classh}(Rj) = \sum_{xp \in classh} \frac{\mu_j(x_p)}{N_{classh}},\ h = 1, 2, \ldots, c \quad (6)$$

where β_{classh} (R_j) is the sum of the compatibility grades of the training patterns in class

R_j is Fuzzy rules

N_{classh} number of training patterns

Step 3: The grade of certainity CF_j is determined as follows

$$CF_j = \frac{\beta_{class\ \hat{h}}(R_j) - \bar{\beta}}{\sum_{h=1}^{c} \beta_{classh}(R_j)} \quad (7)$$

where

$$\bar{\beta} = \frac{\sum_{h \neq \hat{h}_j} \beta_{classh}(R_j)}{(C-1)} \quad (8)$$

The certainty grade for any combination of antecedent fuzzy sets can be specified. Combinations of antecedent fuzzy sets for generating a rule set with high classification ability are to be generated by the fuzzy classification system. When a rule set is given, an input pattern is classified by a single rule as given below.

$$\mu_j(x_p).CF_j = \max\left\{\mu_j(x_p).CF_j | R_j\right\} \quad (9)$$

The winner rule has the maximum product of the compatibility and certainty grade CF_j.

4 Experimentation and Results:

4.1Performance Metric:

The final performance of a classifier can be evaluated by constructing a 2X2 matrix named confusion matrix using the predicted values of the classifier. This matrix will contain both the actual labeled values and predicted class labels. Confusion matrix shows classifications and predicted. The prediction accuracy, also known as classification accuracy, can be calculated from the values of the constructed confusion matrix for a classification problem using the formula given in equation 10.From the confusion matrix the sensitivity and specificity of a classifier also can be calculated. The performance of a classification system should be evaluated by keeping all the instances of the database as a test set. In this work, the splitting of the dataset is done using k-fold cross validation. The dataset is divided into 10 equal partitions and 1 partition out of 10 will be kept as test set and the remaining instances will be used for training the classifier. So the classifier will give ten classification accuracy values for each fold. All the 10 classification accuracy values are listed in Table 2. The average accuracy is calculated from the 10-fold accuracies and is compared with that of the existing systems as shown in Table 3.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (10)$$

4.2 Simulation and Results:

The proposed system was implemented using MATLAB and Weka machine learning software. The weka tool is used for preprocessing and converting the dataset into .arff (Attribute-Relation File Format) format. The values of the data are normalized from 0.0 to 1.0 using the min-max normalization using equation 1.This is done because the values of the fuzzy membership functions ranges from 0 to 1.Now this 0 to 1 range can be used for defining fuzzy membership function of the linguistic variables. The .arff format can represent the attributes in nominal, ordinal or class format. In order to get optimal performance it is important to differentiate between nominal and ordinal attributes. For a single rule the proposed ACO system was run for 200 times. Usually in rule based systems some minimum quantity of data will be left for extracting new rule. This will be called as stop condition value. The stop condition value used here is 0.01.The heuristic rule evaluation function used here is the standard AntMiner+ heuristic (Martens *et al*., 2007).Only one fixed heuristic evaluation function is used in this work with a parametric value of 0.44. A total number of 1000 ants were used for each iteration. The weka along with the AntMiner+ can perform feature subset selection by using the correlation based feature subset evaluation (cfs) method. However this methodology is useful for dataset with more number of attributes that is more than ten attributes. Because this method selects a minimum of 10 attributes. The weka software is also used to split the data into training set and testing set. As we discussed earlier the instances of the dataset are divided into 10-folds using k-fold cross validation technique, to make sure that all the instances in the datasets are trained and tested. Table 2 gives the accuracies for all the 10-folds of cross validation and the average accuracy of the system. The comparison of the performance of the proposed system with other existing systems in the literature is listed in Table 3. From Table 3 it is found that the proposed system performs better than the existing methodologies for this breast cancer dataset by achieving an average accuracy of 95.72%.

Table 2: The accuracy of the system for 10-folds

Fold no.	1	2	3	4	5	6	7	8	9	10	Avg.
Accuracy (%)	95.04	95.91	95.91	95.33	95.33	96.64	95.91	95.91	95.33	95.91	95.72

Table 3: Comparison of accuracy with existing system

Breast Cancer	GA based Approach (Martens *et al*.)	Grid Algorithm (Martens *et al*.)	MLPNN (Nauck and Kruse)	RS+SVM (Ganji and Abadeh)	AR2+NN (Dorigo *et al*.)	**Proposed Fuzzy-ACO**
Accuracy%	94.23	90.78	91.92	96.13	95.6	**95.72**

Conclusion:

In this paper, a clinical decision support system for the breast cancer diagnosis using ACO and fuzzy inference system is proposed. The ACO algorithm along with fuzzy logic was used for extracting rules from dataset and then to optimize them. The fuzzy inference engine was used to perform classification of test data using the optimized set of rules. The proposed system produces an average classification accuracy of 95.72% out of 10-fold cross validation. In future this system can be validated using more standard medical dataset and instead of fixed heuristic evaluation function, more number of heuristic functions can be used.

REFERENCES

Dorigo, M., V. Maniezzo, A. Colorni and V. Maniezzo, 1991. Positive feedback as a search strategy.

Dorigo, M., V. Maniezzo and A. Colorni, 1996. Ant system: optimization by a colony of cooperating agents. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 26(1): 29-41.

Dorigo, M., and C. Blum, 2005. Ant colony optimization theory: A survey. Theoretical computer science, 344(2): 243-278.

Ganji, M.F., and M.S. Abadeh, 2011. A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis. Expert Systems with Applications, 38(12): 14650-14659.

http://onlinelibrary.wiley.com/doi/10.3322/canjclin.45.5.263/pdf, last accessed August 2014.

https://archive.ics.uci.edu/ml/datasets.html, last accessed August 2014.

Huang, C.L., and C.J. Wang, 2006. A GA-based feature selection and parameters optimizationfor support vector machines. Expert Systems with applications, 31(2): 231-240.

Karabatak, M., and M.C. Ince, 2009. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications, 36(2): 3465-3469.

Martens, D., M. De Backer, R. Haesen, J. Vanthienen, M. Snoeck and B. Baesens, 2007. Classification with ant colony optimization. Evolutionary Computation, IEEE Transactions on, 11(5): 651-665.

Nauck, D. and R. Kruse, 1999. Obtaining interpretable fuzzy classification rules from medical data. Artificial intelligence in medicine, 16(2): 149-169.

Polat, Kemal, and Salih Guneș., 2007. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing, 17(4): 694-701.

Polat, K., and S. Guneş, 2007. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing, 17(4): 694-701.

Ubeyli, E.D., 2007. Implementing automated diagnostic systems for breast cancer detection. Expert Systems with Applications, 33(4): 1054-1062.