Appl Intell (2012) 37:357–376
DOI 10.1007/s10489-011-0332-x

Optimizing the modified fuzzy ant-miner for efficient medical diagnosis

Thannob Aribarg · Siriporn Supratid · Chidchanok Lursinsap

Published online: 6 January 2012

Abstract The advantage of efficient searches belonging to ant-miner over several other approaches leads to prominent achievements on rules mining. Fuzzy ant-miner, an extension of the ant-miner provides a fuzzy mining framework for the automatic extraction of fuzzy rules from labeled numerical data. However, it is easily trapped in local optimal, especially when it applies to medical cases, where real world accuracy is elusive; and the interpretation and integration of medical knowledge is necessary. In order to relieve such a local optimal difficulty, this paper proposes OMFAM which applies simulated annealing to optimize fuzzy set parameters associated with a modified fuzzy ant-miner (MFAM). MFAM employs attributes and training case weighting. The proposed method, OMFAM was experimented with six critical medical cases for developing efficient medical diagnosis systems. The performance measurement relates to accuracy as well as interpretability of the mined rules. The performance of the OMFAM is compared with such references as MFAM, fuzzy ant-miner (FAM), and other classification methods. At last, it indicates the superiority of the OMFAM algorithm over the others.

T. Aribarg · S. Supratid (✉)
Department of Information Technology, Rangsit University, Pathumthani, 12000, Thailand
e-mail: siri_sup1@hotmail.com

T. Aribarg
e-mail: iamaribarg@yahoo.com

C. Lursinsap
Advanced Virtual and Intelligent Computing Lab, Department of Mathematics, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand
e-mail: lchidcha@gmail.com

Keywords Ant-miner · Fuzzy logic · Simulated annealing · Adaptive neuro-fuzzy inference system · Multi-class support vector machine

1 Introduction

Early detection of life applications regarding medical problems, is important to increase the chance of successful treatment. Such detection is often formulated as a classification problem. The aim of the classification task is to assign the sample cases to related classes, out of a set of predefined classes, based on the values of some attributes for the cases [1]. One of the successful methods is multi-class support vector machine ($SVM^{multiclass}$) [2]. Such a Multi-class SVM generates a hyperplane to separate several classes of data sets. It derives a class decision by determining the separate boundary with maximum distance to the closest points, namely support vectors, of the training dataset. Similar to the weights derived by neural learning, such extracted support vectors represent a cryptic form of knowledge; therefore, are incomprehensible. $SVM^{multiclass}$ is also used in such medical applications [3–5]. Fuzzy approaches have become one of the well-known solutions for classification problems. Fuzzy logic [6] improves classification [7] and decision support systems by allowing the powerful use of overlapping class definitions. This efficiently handles uncertainty and vagueness, especially consisted in medical diagnosis applications [8]. Furthermore, the use of fuzzy IF-THEN rules is represented in linguistic forms that are easily interpreted and examined by humans [9]. Not only improving the interpretability of the classification results, fuzzy rules also provide more insight into the classifier structure and decision making process [10, 11]. A sample case can be assigned to several classes with different degrees of membership. A well-known adaptive neuro-fuzzy inference system (ANFIS) [12] is a specific approach that combines neural learning and fuzzy systems. The striking characteristics of ANFIS are its ability to represent fuzzy rules in a form of IF-THEN that remedies incomprehensibility problems, as well as its hybrid learning of the gradient method, and the least squares estimate (LSE). Successful implementations of ANFIS in several data mining researches including biomedical areas have been reported [13, 14]. However, some disadvantages of ANFIS such as local optimal trap emerged by gradient descent method are unavoidable; besides, it is also tediously complicated to make modification or variation apart from both gradient descent and LSE learning, since they are embedded in the network. In addition, the works belonging to [15, 16] also produced fuzzy IF-THEN rules. A GA-based algorithm was introduced for selecting a small number of significant fuzzy rules from a larger candidate rule set. The aim was to construct a fuzzy classification system with high accuracy. Later, a powerful fuzzy modeling scheme was proposed in [17] for complexity reduction. The multiple GA-based algorithms were used to produce fuzzy rules [18–21]. They applied Ant Colony Optimization (ACO) algorithm as a local search procedure and improved the performance of their final classification system. ACO introduced by [22], is a heuristic algorithm with efficient local search for combinatorial problems [23–25]. ACO imitates the behavior of real ant colonies in nature to search for food and to connect to one another by pheromones laid on paths traveled [26]. This algorithm has been developed significantly as a probabilistic search algorithm for large scale optimization problems, and its use arises frequently in real applications [23]. The research [27] introduced ACO to classification task for the first time and called it ANTMINER. The ANTMINER deployed artificial ants to construct a set of classification rules. The results were promising. A simpler function instead of entropy was used in [28] as heuristic information in order to reduce the computation overhead. They called their algorithm ANTMINER2. Also, [29] introduced another version of ANTMINER (named ANTMINER3). More exploration was encouraged by means of a different transition rule in this version. After that, ANTMINER+ was presented by [30]. It was a classification technique based on a *MAX-MIN* ant system. ANTMINER+ achieved an average accuracy that was significantly better than the previous ANTMINER, ANTMINER2 and ANTMINER3 in most of the data sets. Fuzzy rules from ANT-inspired computation—simultaneous rule learning (FRANTIC-SRL) was presented by [31]. It used a concept of ant-miner for mining fuzzy rules. Several fuzzy ant-miner algorithm instances were executed in parallel, where each instance generated rules for a particular class. Fc-AntMiner was proposed by [32]. Artificial ants were employed to explore the search space and gradually construct candidate fuzzy rules. The algorithm concerned with balancing the cooperation and competition between the ants. This was done so that the ants were encouraged to find more accurate rules. A correlation-based ant miner (AntMiner-C) [33] was proposed for classification rule discovery. Its main feature related to the use of a heuristic function based on the correlation between the recently added term, and the term to be added in the rule. In addition, the weight of a training case and that of an attribute were introduced in FCACO, which represented the fuzzy classification rules mining algorithm with ACO [34]. Such weights were applied for improving classification accuracy. FCACO demonstrated the integration of ACO for search strategy, and fuzzy set for representation of the rule terms to cope with continuous values. The modification of fuzzy ant-miner (FAM), based on the aforementioned weight of a training case, and that of an attribute gives motivation to our research. Such a modified FAM is named here, MFAM.

Importantly, the considerable factor, which usually causes local optimal in the fuzzy ant-miner instances, comes from the lack of a structural method for efficiently confirming the competent selection of fuzzy parameters. Simulated Annealing (SA), introduced by [35] is a global optimization method that distinguishes between different local optima. It provides a framework for optimization of the parameters of very large and complex systems. By using SA for optimizing the selection of fuzzy set parameters, powerful classification would be granted to MFAM.

This paper proposes a such powerful classification method, using SA for optimizing the modified fuzzy ant-miner. Such proposed method is named here OMFAM. OMFAM, optimized by SA has a capability to avoid local optima, and attain global optimal solution. Through SA optimization, OMFAM is brought to achieve better quality, and more accurate rule results. OMFAM is applied with six critical medical diagnosis data sets from UCI repository [36], for assessing the classification mining methods. The rest of this paper is organized as follows. In Sect. 2, related works on classification mining methods as well as their effectiveness are mentioned. The detail of MFAM is explained in Sect. 3. Then, OMFAM is explained in Sect. 4. The results are shown in Sect. 5. In this section, the performance measurement according to rules accuracy as well as interpretability is declared; and the comparison study is done among OMFAM and the others including MFAM, FAM, SVM and ANFIS. Discussion section concerns with the description of OMFAM accomplishment compared to other authors' work. Finally, the overall conclusion is drawn in the last section to show the credibility of using SA optimization regarding OMFAM.

2 Related works

There have been a lot of studies, reported in the literature in which the researchers have used medical data sets to evaluate their classification works. First, Wisconsin breast cancer dataset was used for classification performance evaluation. ANTMINER was an algorithm for data mining. 93.84% accuracy was reported for such cancer data set [27]. The accuracy of 97.17% was obtained in [37] with the use of the mutual correlation-based feature selection and fuzzy k-nearest neighbor (KNN). ANTMINER+ completed 96.40% [30]. Such accurateness was competitive or even better than SVM, Neural Network, C4.5, KNN and Naive Bayes. FCACO [34] and Fc-AntMiner [32] were employed to mine fuzzy rule set. They sequentially achieved 95.26% and 97.51% classification accuracy. AntMiner-C was also assessed; and achieved 97.54% accurate rate, using the breast cancer data [33]. Another medical diagnosis data, thyroid gland was utilized for estimating the quality of classification as well. The Probabilistic Neural Network (PNN) performed a comparative study on an thyroid dataset [38]. The result shows 94.81% classification accuracy. In [39], the fuzzy rules were found by Expert System for Thyroid Disease Diagnosis (ESTDD), using neuro-fuzzy method. Thyroid diseases could be detected with 95.33% accuracy by the expert system. GDA_WSVM represented a generalized discriminant analysis and wavelet support vector machine system [40]. 91.86% classification accuracy was obtained for diagnosing thyroid disease. SIM, a fuzzy similar model was another classification method that was evaluated using thyroid gland data set [41]. It applied generalized mean to the thyroid data; and received 96.86% accuracy. Echocardiogram dataset was also employed to estimate the quality of classification mining methods. TACO-miner obtained 96.40% accuracy for echocardiogram data by using a rule extraction from neural networks, via ant colony algorithm for data mining applications [42]. The classification method proposed a rule extraction from trained adaptive neural networks using artificial immune systems (AIS) [43]. The accuracy of 94.59% was obtained for echocardiogram dataset. FDSVM [44] used support vector machine incorporated with feature discrimination. Classification accuracy of 87.69% was achieved for the same data.

3 The modified fuzzy ant-miner (MFAM)

A traditional ant-miner, ANTMINER is applied for mining classification crisp rules. The rules are generated based on a set of N training cases $Tr = \{(\mathbf{x}^1, c^1), \ldots, (\mathbf{x}^n, c^n), \ldots, (\mathbf{x}^N, c^N)\}$, where $\mathbf{x}^n = \{x_1^n, \ldots, x_d^n, \ldots, x_D^n\}$; x_D^n specifies attribute x_D of case n; c^n is a class label corresponding to $\mathbf{x}^n$. Fuzzy ant-miner (FAM) [31], an extension of the ant-miner generates fuzzy rule mining instead of crisp rule mining. R^i is the label of the ith fuzzy if-then rule, constructed by FAM.

$$R^i : \text{IF } x_1 \text{ is } A_{1j} \text{ and } \ldots x_d \text{ is } A_{dj} \ \ldots \text{ and } x_D \text{ is } A_{Dj}$$
$$\text{THEN } class = c^n$$

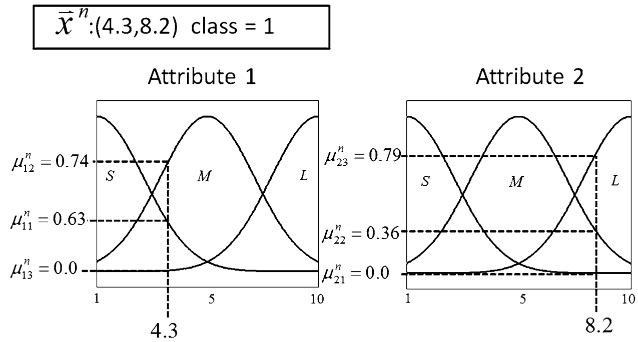


Fig. 1 Data fuzzification

where 'x_1 is A_{1j} and $\ldots x_d$, is $A_{dj} \ldots$ and x_D is A_{Dj}' refers to the condition or antecedent part; a single term 'x_d is A_{dj}' could be represented by $term_{dj}$; '$class = c^n$' is specified in the consequent part. In accordance with fuzzy classification rule, $term_{dj}$ is referenced by a linguistic term. The possible linguistic $term_{dj}$ here is defined as one of the following: 'x_d is S', 'x_d is M' or 'x_d is L', where S, M and L abbreviate for '*Small*', '*Medium*' and '*Large*' sequentially. Such three linguistic values are delineated by fuzzy sets on the unit interval [0, 1] as shown in Fig. 1. The membership function of each linguistic value in Fig. 1 is specified by homogeneously partitioning the domain of each value of $term_{dj}$ into Gaussian function. Figure 1 gives an example of degree of membership or degree of matching between the value of $term_{dj}$ and Gaussian fuzzy sets S, M and L in the condition part. In Fig. 1, case is assumed to contain 2 attributes, (4.3, 8.2). The value 4.3 of first term is fuzzified into fuzzy sets S, M and L with membership or matching degree $\mu_{11}^n \approx 0.63$, $\mu_{12}^n \approx 0.74$ and $\mu_{13}^n \approx 0.0$ consecutively; likewise, the value 8.2 of the second term is also fuzzified into the same sets of fuzzy function with membership of $\mu_{21}^n \approx 0.0$, $\mu_{22}^n \approx 0.36$ and $\mu_{23}^n \approx 0.79$.

However, according to FAM, the sample cases that are correctly classified by or matched with the mined rule are removed from the training set. Rules mined in later stages are unaware of the previously removed cases and therefore might be in conflict with rules mined earlier. Unexpected interactions between rules can appear when a case is covered by several rules of different classes. In order to relieve such conflict problem, the modified fuzzy ant-miner (MFAM) is presented. MFAM employs two concepts in FCACO, which are the training case, weighing, and attributes weighing. These two types of weighing concepts are applied in the fuzzy rule construction process. Unlike FAM that immediately cut off the training case that matches the best constructed rule, MFAM utilizes the training case weight, $w_{\mathbf{x}^n}$. Instead of instantly removing the matched case $\mathbf{x}^n$, in MFAM the weight of that case is reduced. By this manner, the rules mined in later stage would be given some chance to become aware of the matched cases, existing in

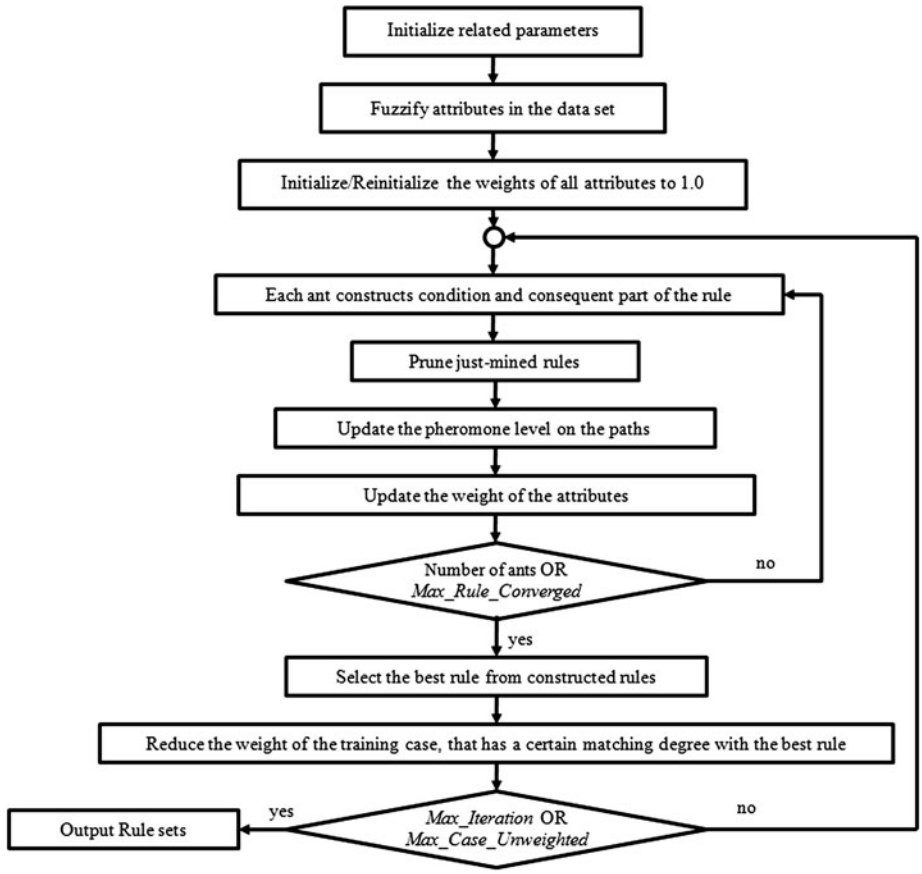


Fig. 2 Framework of an MFAM

previous stage. This lessens the conflict between the current mined rule and the rules mined earlier. The other concept is denoted as the weight of $term_{dj}$, which is referenced as $wTerm_{dj}$. In MFAM, $wTerm_{dj}$ contained in the mined rule is reduced whenever the rule is defined as the best rule, and is inserted into rule list. The purpose of such weight reduction is for avoiding from mining the same rule. Thus, the unexplored rules would have better chance to be mined.

Figure 2 displays the steps of rule construction process in MFAM, which is similar to those in classic ANTMINER. At the beginning, the related parameters are initialized. The number of ants is assigned; and the rule sets are initialized to empty. In MFAM, the weights of all training cases as well as the weights of all attributes are set to 1.0. Each ant, while it is walking, deposits a chemical substance on the passed $term_{dj}$ called pheromone, τ_{dj}. Such pheromone encourages the following ants to stay close to the previous best ant. All single $term_{dj}$ in the condition part are initialized with the same amount of pheromone. The initialized amount of pheromone is inversely proportional to the number of values of all attributes. This means that all terms of attributes in the condition part have the same probability of being chosen by an ant. After initialization, the value of each $term_{dj}$, contained in each training case is fuzzified; and the membership degree of each $term_{dj}$ with fuzzy set S, M and L is obtained. Then, each ant constructs the condition and consequent part of a single rule. A condition part is constructed by selecting $term_{dj}$ with local best quality. An amount of probability P_{dj}, computed mainly influents the selection of the local best quality term. P_{dj} is computed by (1). Such probability relies on the normalization of the amount of pheromone τ_{dj} and a problem dependent heuristic value η_{dj}, associated with each $term_{dj}$.

$$P_{dj} = \frac{\tau_{dj}\eta_{dj}}{\sum_{d=1}^{D}\sum_{j=1}^{J_d}\tau_{dj}\eta_{dj}} \tag{1}$$

Heuristic value, η_{dj} is taken to be an information theoretic measure for quality of $term_{dj}$ to be selected and added to rule condition. The heuristic value is measured by the following (2):

$$\eta_{dj} = \frac{\log_M M - entropy(term_{ij})}{\sum_{\forall d}\sum_{\forall j}(\log_M M - entropy(term_{ij}))} \tag{2}$$

where M is the number of possible classes. The quality of $term_{dj}$ is measured with regard to the entropy ($term_{dj}$) or amount of information of $term_{dj}$. Such entropy is defined in (3).

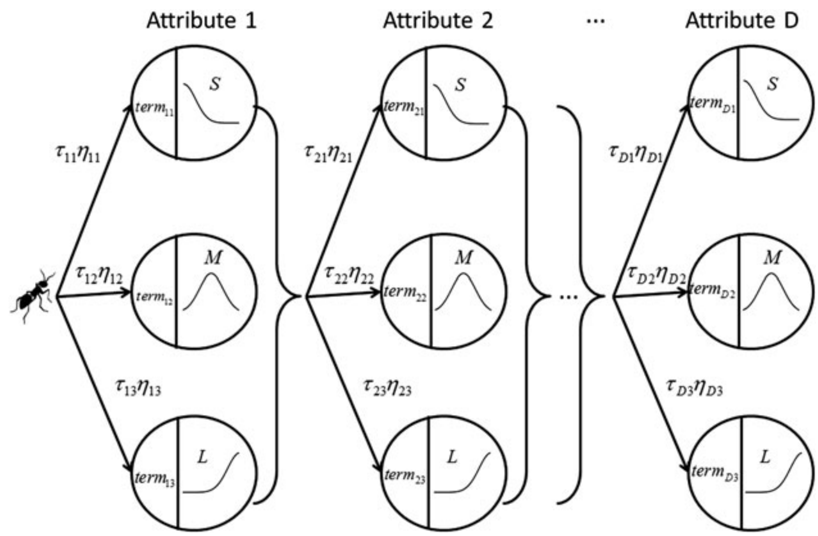


Fig. 3 A visual view of rule construction according to MFAM where S, M and L are linguistic values associated with each $term_{dj}$

$$entropy(term_{dj}) = -\sum_{m=1}^{M}\left(\frac{AllCases\_WtermDM\_C^m}{AllCases\_WtermDM}\right) \times \log_M\left(\frac{AllCases\_WtermDM\_C^m}{AllCases\_WtermDM}\right) \quad (3)$$

The entropy of $term_{dj}$ is represented by entropy $(term_{dj})$ for all classes. $AllCases\_WtermDM\_C^m$ is identified by summation of degree of matching between the value of $term_{dj}$ and fuzzy set j for all cases in class c^m, where j is specified as one of the fuzzy sets S, M or L) in the rule, being constructed. $AllCases\_WtermDM$ has similar definition as $AllCases\_WtermDM\_C^m$, except all cases for entire number of classes are taken into account. However, the proportion $AllCases\_WtermDM\_C^m$ to $AllCases\_WtermDM$ is normalized by log base M of such proportion itself. It is also notable that the degree of matching is computed with some weight. That weight is equivalent to weight of $term_{dj}$, $wTerm_{dj}$. Such weight is taken into account with entropy for avoiding mining the same rule. The weight is reduced by 1% whenever the rule is inserted to the rule list. Considering (2) and (3), the higher the entropy, the less the heuristic value is. The high entropy as well as the less heuristic value signifies the high degree of uniformly distributed to the classes and the low classifying power of $term_{dj}$, and vice versa. Therefore, selecting $term_{dj}$ with the highest heuristic value, η_{dj} would improve the quality of rule classification. As mentioned before, the pheromone τ_{dj} encourages the following ants to stay close to previous ants. It means the previous ant suggests the following ant the best $term_{dj}$ to select. At this point, one would say the higher amount of pheromone and heuristic value, the higher probability P_{dj}, that the $term_{dj}$ would be selected. However, such manner of term selection tends to exploitation, that possibly leads to local optima. A certain regulation may be set up for the sake of balancing between the exploration and exploitation. The highest probability is chosen to use in term selection, if exploitation P_{dj} is decided. The average of P_{dj} for all related fuzzy sets j would be used, if the decision is made on exploration. The choice between those two depends on a specific uniform random number.

Figure 3 delineates a visual view of constructing the condition part of a rule. Each node represents an individual $term_{dj}$; possibly selected by the ant. For each $term_{dj}$, one of the possible values of fuzzy set S, M and L is to be selected. A certain $term_{dj}$ is selected and added into the condition part one-term-at-a-time.

Following the construction of the condition part of the rule, the consequent part is determined to complete the rule construction. The consequent class is decided by calculating the matching degree of each training cases $\mathbf{x}^n$ with the fuzzy rule R^i. The calculation is done by the product operation: $\mu^n_{R^i} = \mu_{R^i_1}(x^n_1) \times \cdots \times \mu_{R^i_d}(x^n_d) \times \cdots \times \mu_{R^i_D}(x^n_D)$. As stated earlier, '$x_d$ is A_{dj}' has corresponded to $term_{dj}$. By such correspondence, $\mu_{R^i_d}(x^n_d)$ is equivalent to matching degree of $term_{dj}$, belonging to case $\mathbf{x}^n$ and the corresponding term, belonged to rule R^i. Each possible classes C_m accumulates the degree of activation of fuzzy rule R^i with a matching consequence $c^i = C_m$. The consequent class to which the case $\mathbf{x}^n$ is classified is the majority class, C^n_{major}. Such decision making is shown in (4).

$$C^n_{major} = \arg\max_{C_m} \sum_{R^i|c^i=C_m} \mu^n_{R^i} \quad (4)$$

Immediately after the ant has completed a rule construction, pruning is undertaken to increase the interpretability and accuracy of the mined rule. Local heuristic function considers only one-term-at-a-time, ignoring term interactions. By this reason, irrelevant terms may have been included in the rule due to stochastic variations in the term selection procedure or due to the use of a shortsighted selection. The basic idea is to iteratively remove one-term-at-a-time from the mined rule, while the pruning process improves the quality of the rule [27]. After the pruning step, the rule may be assigned a different consequent class based on the majority class in the cases, covered by the rule condition. The rule pruning procedure iteratively removes the term whose removal will cause a maximum increase in the quality of the rule. The quality of a mined rule, Q is measured by (5)

$$Q = Q_1 + 0.005 Q_2 \tag{5}$$

From (5), the quality Q of the rule relies on 100% of correctness estimation, represented by Q_1 and 0.5% of interpretability, represented by Q_2. The correctness of the rule, Q_1 depends on sensitivity, specificity along with 1% of accuracy. Sensitivity relates to the accuracy among positive samples; specificity concerns the accuracy among negative samples. Accuracy indicates all correct results in the population. Q_1 is defined in (6)

$$\frac{sensitivity \times specificity + 0.01 \times accuracy}{1 + 0.01} \tag{6}$$

where $sensitivity = \frac{TP}{TP+FN}$, $specificity = \frac{TN}{TN+FP}$ and $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$. In case of crisp data mining, sensitivity denotes the proportion of cases covered by the mined rule, having the same class as that rule (true positive) to all cases, having the same class as that rule (all of true positive and false negative). On the other side, specificity is defined as the proportion of cases that are not covered by the mined rule, having a different class from that rule (true negative) to all cases, having a different class from that rule (all true negative and false positive). Accuracy rate is defined by the proportion of the total of cases covered by the mined rule, having the same class as that rule (true positive) and cases that are not covered by the mined rule, having a different class from that rule (true negative) to the whole number of cases (*TP*, *TN*, *FP* and *FN*). Instead of crisp rules mining rules, the fuzzy mining is taken into account in MFAM. A case can be covered by a rule condition to a membership degree of that case in that rule condition. The sensitivity, specificity and accuracy computed in fuzzy rules mining are based on the same idea as what is done in crisp rules mining one. The denotation of sensitivity, specificity and accuracy are consecutively illustrated in (7)

$$\begin{aligned} sensitivity &= \frac{AllCases\_WcaseDM\_C^R}{AllCases\_C^R} specificity \\ &= \frac{AllCases\_NotWcaseDM\_NotC^R}{AllCases\_NotC^R} \\ accuracy \\ &= \frac{AllCases\_WcaseDM\_C^R AllCases\_NotWcaseDM\_NotC^R}{AllCases} \end{aligned} \tag{7}$$

$AllCases\_WcaseDM\_C^R$ is identified by summation of degree of matching between case $\mathbf{x}^n$ and the mined rule R, having the same class as rule R. $AllCases\_NotWcaseDM\_NotC^R$ is denoted as summation of negated degree of matching between case $\mathbf{x}^n$ and the mined rule R, having a different class from rule R. $AllCases\_C^R$ refers to all cases, having the same class as rule R. Vice versa, $AllCases\_NotC^R$ is all cases, having a different class from rule R. In traditional ANTMINER, the matched cases in previous-stage has not had the opportunity to get into account of determining quality of rule R in later-stage. This leads to the conflict between the current mined rule and the rules mined earlier. Due to such difficulties, the weight of training case $\mathbf{x}^n$, $w_{\mathbf{x}^n}$ would be used. The weight $w_{\mathbf{x}^n}$ where $\mathbf{x}^n$ matches the best rule, with matching degree $\mu^n_{R^{best}} \geq 0.8$ would be reduced by an amount of matching degree of $\mathbf{x}^n$ and the best rule, $\mu^n_{R^{best}}$, $w_{\mathbf{x}^n}(new) = w_{\mathbf{x}^n}(old)(1 - \mu^n_{R^{best}})$. By such a reduction manner, one would say the stronger the matching degree is, the more reduction rate the weight $w_{\mathbf{x}^n}$ would have. This would give little chance for a strongly matched case to exist in the later stage; would give rather moderate or high chance for a weakly matched case to do so. However, there is a certain time that the training case $\mathbf{x}^n$ is absolutely cut off. That is when the corresponding $w_{\mathbf{x}^n}$ is less than a threshold 0.2.

Apart from the correctness of the mined rules measured by Q_1, the rules interpretability is measured by Q_2. Here the quality, concerning with length of the rule is approximated by (8).

$$Q_2 = 1 - \frac{NumberOfTerms}{NumberOfAllTerms} \tag{8}$$

From (8), *NumberOfTerms* denotes number of terms, used in the mined rule; and *NumberOfAllTerms* is all possible number of all terms. If the ratio of *NumberOfTerms* to *NumberOfAllTerms* is small, then the small length of the rule is indicated, and vice versa. This points out the conciseness of the rule. A concise rule denotes the good quality Q_2.

After the ant has completed a one rule construction and rule pruning, the pheromone is accumulated for all $term_{dj}$ in the condition part of the just-mined rule; and is evaporated in unused term. The highest amount of pheromone points to the $term_{dj}$ of the rule, constructed by the previous ant. The high amount of pheromone would encourage the following ants to stay close to the previous ant. This possibly leads to the better quality mined rule in the next generation of ant colony. The pheromone accumulation depends on the quality of the just-constructed rule as shown in (9)

$$\tau_{dj}(new) = \tau_{dj}(old) + \tau_{dj}(old) * Q \tag{9}$$

On the other side, the pheromone evaporation is achieved by dividing the value of pheromone on each unused term by the summation of the pheromone in all terms. The end of rule construction is accomplished by a single ant at this point.

Then, the next ant would repeat the rule construction process. Such repetition is processed until one of the following two criteria is met. First, a predefined *Max_No_Ants* is attained, that means all ants in the colony accomplish the rule construction. Second criteria is when the constructed rules have not changed throughout a predefined value, *Max_Rule_Converged*. After all ants accomplish on mining rules, the best rule with the highest quality Q is selected; and is added to the best rule list.

The moment that one ant colony completes rule construction, the next colony of ants is going to iterate the process of rule mining. The parameters such as $wTerm_{dj}$ and $w_{\mathbf{x}^n}$ are initialized to 1.0. $wTerm_{dj}$ is reset for allowing the new colony of ants to select condition terms in a full search space. $w_{\mathbf{x}^n}$ is reset as well to permit a full chance for all cases to be counted in rule construction process. The rule mining process is iterated until one of these three events occur: one is when the number of cases in the training set, that are left unweighted is less than *Max_Case_Unweighted*; another one, when the current ant has constructed a rule that is exactly the same as the rule constructed by the two previous ants; the last one is when *Max_Iterations* is met. With a final set of the mined rules, fuzzy rule inference can be done for a new unknown case, $\mathbf{x}^{new}$. The consequent class of that $\mathbf{x}^{new}$ would be predicted by such inference just like that, predicted in the consequent part or rule construction.

It is noticeable that the fuzzy sets, used for rule construction in MFAM are fixed through the whole process. The fixed fuzzy sets are delineated based on Gaussian membership function, a universal approximator. Generically, the underlying data is modeled by such Gaussian function. The knowledge about data set to be fuzzified is necessary in order to define parameters appropriately; it applies simply generic parameters such as mean and standard deviation, based on attributes of such data set. Nevertheless, it may be possible that setting fuzzy sets in such manner yields unsatisfied results. The fixed parameters of all fuzzy sets may not be suitable for several situations. Inappropriate fuzzy set parameters would lead to local optimal trap. The proper selection of means as well as standard deviations for fuzzy set function is important to the accuracy of MFAM classification. In accordance with MFAM, structural methods for efficiently confirming the effective selection of parameters are lacking. Simulated Annealing (SA) is a generic probabilistic metaheuristic for global optimization problem of locating a good approximation to the global optimum of a given function in a large search space [35]. To confirm the productive selection of fuzzy set parameters, the SA is used to globally optimize such parameters selection in the proposed mining model. The optimization of MFAM, using SA is described in the next section.

4 Optimizing the modified fuzzy ant-miner by using simulated annealing (OMFAM)

As aforementioned, the quality of rule construction in MFAM may be destroyed by improper fuzzy set parameters that are fixed through the whole execution. In order to improve MFAM to avoid from local optimal problem, the proposed method utilizes Simulated Annealing (SA) for dynamically finding the optimized fuzzy set parameters within MFAM. Such a method is called here, OMFAM. SA is an optimization technique, analogous to the annealing process of material physics. Boltzmann pointed out [45] if the system is in thermal equilibrium at a temperature T, then the probability $P_B(S)$ of the system being in a given state s is given by the Boltzmann distribution:

$$P_B(s) = \frac{\exp(-E(s)/kT)}{\sum_{w \in S} \exp(-E(w)/kT)} \tag{10}$$

where $E(s)$ denotes the energy or fitness of state s; and S is the set of all possible states. However, (10) does not contain information on how a fluid reaches thermal equilibrium at a given temperature. Metropolis algorithm [46] is developed for simulating the process of Boltzmann. The Metropolis algorithm is summarized as follows. When the system is in original state old with energy $E(s_{old})$ a randomly selected atom is perturbed, resulting in a state s_{new} with energy $E(s_{new})$. This new state is either accepted or rejected depending on the Metropolis criterion: if $E(s_{new}) \leq E(s_{old})$ then the new state is automatically accepted, in contrast, if $E(s_{new}) > E(s_{old})$ then the probability of accepting the new state is given by the following probability function.

$$P(\text{accept}\, s_{new}) = \exp\left(-\frac{E(s_{old}) - E(s_{new})}{T}\right) \tag{11}$$

The Metropolis approach is conducted for each temperature on the annealing schedule until thermal equilibrium is reached. Additionally, a prerequisite for applying SA is that a given set of the multiple variables defines a unique system state, for which the objective function can be calculated. The SA algorithm which is applied for optimizing MFAM is described as follows.

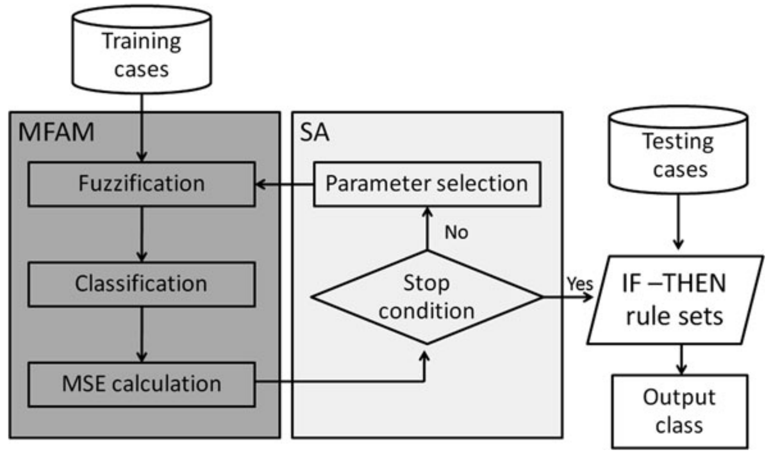
Step 1: Initialization

Here, a Gaussian membership function G_{dj} is utilized as a fuzzy set, S, M, L.

$$G_{dj} = \exp\left\{-\left(\frac{x_j - m_{dj}}{v_{dj}}\right)^2\right\} \tag{12}$$

where m_{dj} and v_{dj} the center (mean) and width (standard deviation) of fuzzy set, associated with $term_{dj}$. The initial value of the fuzzy set parameters is generated by center and width, based on attributes of such data set, as mentioned before; and they are fed into MFAM model in a following form.

Fig. 4 Framework of an OMFAM



$$s = [m_{1S}, m_{1M}, m_{1L}, v_{1S}, v_{1M}, v_{1L}, \ldots, m_{dS}, m_{dM}, m_{dL}, v_{dS}, v_{dM}, v_{dL}, \ldots, m_{DS}, m_{DM}, m_{DL}, v_{DS}, v_{DM}, v_{DL}] \quad (13)$$

The positive classification error is defined as $E(s)$, the fitness of state s. Here, the initial state s is equivalent to initial fuzzy set parameters.

Step 2: Calculate new state, s_{new}

A new state, s_{new} is a solution nearby the current state, s_{old}; it is generated by a few random moves from the old state in a certain range. The ranges of $[-0.5, 0.5]$ and $[-0.1, 0.1]$ respectively applied for the move of center (m_{dj}) and width (v_{dj}) of s_{old}. Such a move of m_{old} and v_{old} consisted in s_{old} is illustrated by (14).

$$\begin{aligned} m_{new} &= m_{old} + Random(-0.5, 0.5) * m_{old} \\ v_{new} &= v_{old} + Random(-0.1, 0.1) * v_{old} \end{aligned} \quad (14)$$

Step 3: Make acceptance tests

To determine the acceptance or rejection of new state, (15) is employed

$$\begin{cases} Accept\ the\ new\ state, & \text{if } E(s_{new}) > E(s_{old}),\ \text{and} \\ & random(0,1) < P(accept\ s_{new}) \\ Accept\ the\ new\ state, & \text{if } E(s_{new}) \le E(s_{old}) \\ \text{Re}ject\ the\ new\ state, & \text{otherwise} \end{cases} \quad (15)$$

If the new state is accepted, then set the new state would be set as the current state.

Step 4 : Find better solutions

If the new state is not accepted, then steps 2 and 3 are repeated until the s_{new} is superior to s_{old}, or a certain criteria is met. Then, s_{new} is set as the current state.

Step 5 : Reduce temperature

After the new system state is obtained, the temperature is reduced by one. If the zero temperature is reached, then the algorithm stops, and the latest state is an approximate optimal solution. Otherwise, go to step 2. In the optimization, the values of the mean of square errors (MSE), shown as (16) serve as the criterion for identifying suitable fuzzy set parameters for OMFAM model

$$\begin{aligned} \text{MSE} &= \frac{\sum_{n=1}^{N} \text{Class\_Error}}{N} \\ \text{Class\_Error} &= \begin{cases} 0, & \text{if } c_n^{\text{Actual}} = c_n \\ 1, & \text{if } c_n^{\text{Actual}} \neq c_n \end{cases} \end{aligned} \quad (16)$$

where N is the number of training cases ; c_n^{Actual} the actual class predicted for training case $\mathbf{x}^n$; and c_n the class predicted by the rule value for the same training case $\mathbf{x}^n$. Figure 4 illustrates the framework of the proposed OMFAM model. The SA algorithm is used to seek a better combination of the fuzzy set parameters in MFAM model. When a new state, s_{new} of SA algorithms is determined, the values of fuzzy parameters in s are evaluated on MFAM. Then, a classification process is conducted, and a classification error, MSE is obtained. Finally, if the zero temperature is attained, then the algorithm stops, and the latest solution, s is an approximate optimal solution. Later on, this optimal solution set of fuzzy parameters would be applied in OMFAM to classify the test or unseen cases.

Although the optimization exists in the mining process, $O(N)$ or $O(Max\_Iterations)$ still represents the worst-case complexity for OMFAM. The SA-related parameters are counted as small value constants.

5 Experimental results

Comparative evaluation is carried among fuzzy ants classification approaches, OMFAM, MFAM and FAM as well as those of effective neural miners such as SVM and ANFIS. The performance assessment is done based on two main criteria: accurateness and interpretability of the classification results of the mined rules. Here, the accuracy of the

Table 1 Parameters used in the mining methods

Parameters	Mining methods	Value
Max_No_Ants	OMFAM, MFAM, FAM	5
Max_Rules_Converge	OMFAM, MFAM, FAM	3
Max_Case_Unweighted	OMFAM, MFAM	5% of training cases
Max_Iterations	MFAM, FAM, ANFIS	200
Max_Iterations × *Max_Temperature*	OMFAM	10 × 20

mined rules is evaluated by average accuracy of classification on the account of training and testing cases as well as overfitting, and yielded by the mining methods. Additionally, sensitivity, specificity and area under the receiver operating characteristics curve (AUC) [47] are reported also, the deviation from the accuracy median is determined; and is displayed in a boxplot graph. The other criterion, interpretability of the rules is estimated by the number of resulted mined rules, number of terms per rule, as well as percentage of attribute terms that are included in the mined rule.

Table 1 defines a set of parameters, associated with the mining methods. Number of iteration runs is one of the factors that affect the effectiveness of mining methods. However, explicit distinction of algorithm structures is notable between the two types of mining approaches, fuzzy ant classification methods, and those concerning neural classifiers. Fuzzy ant-miner approaches apply multi-agent concepts to achieve classified mined rules. With respect to these ant miner methods, the termination of iterative runs depends upon any of the following factors, predefined maximum number of iterations; represented by *Max_Iterations*, maximum number of ants, number of rule convergence tests, and maximum number of remaining unweighted cases in the training set as well. The maximum number of ants, referenced by *Max_no_Ants* in Table 1 denotes the number of ants at maximum, employed within rule construction in each iterative run. Assigning too many ants would consume too much runtime, however, too small number of them may not produce good results. The next two parameters, that have influence on the number of iterative runs are the number of rule convergence tests and maximum number of remaining unweighted cases in the training set respectively represented by *Max_Rules_Converege* and *Max_case_Unweighted*. The first parameter is assigned the largest number of repetition of mined rules, allowed in a session of an ant run. An appropriate parameter assignment lessens a chance to produce the repeated mined rules. The other one, *Max_case_Unweighted* represents the threshold number of remaining unweighted training cases allowed for termination. Such unweighted training cases indicate the cases that have never matched any constructed rules. Too high value of the threshold may reduce opportunity to explore the new or better rules; too low of that may cause rule overfitting. Besides the fuzzy ant-mining approaches, the other type of mining methods, referred in Table 1 are effective neural-based classifiers. This type of classifier mines data by finding the optimized model of weight set. The weight set model yields minimum learning error less than a very small amount of threshold. ANFIS optimizes such neural weights to achieve the mined best rules; whilst SVM uses the optimal set of support vectors to do the same things. Table 1 points out the same maximum iterations, *Max_Iterations* which is consumed by the related mining methods, except OMFAM. A few numbers of ants, used by MFAM and FAM gives very little effect, compared to 200 maximum iterations. The 10 maximum iterations is run by MFAM part, consisted in OMFAM. In addition, 20 maximum iterations of SA is also counted in OMFAM. Therefore, the total of 200 maximum runs is regulated with respect to OMFAM. Thereby, one would say all comparative mining methods have about the same maximum iteration runs, nevertheless some method may get converged before the specified maximum iterations, by the reason earlier described.

To evaluate all related mining methods, six benchmark medical data sets obtained from the UCI repository databases [36] are considered. The major characteristics of those data sets are summarized in Table 2.

From Table 2, Among the related data sets, the number of attribute (#Attri.) terms varies in rather wide range. The real and nominal types of attribute term are pointed out. The number of cases, scattered in each class is implied by the deviation of class distribution which is represented by Dev.Cla. in Table 2. One would see the explicit difference of such deviation between the first three data sets and those of latter three. Such difference relates to the distinction between the percentages of majority class (Maj.Cla.) and those of minority class sample cases (Min.Cla.). For each dataset, the entire data is randomized; and then is used to prepare ten sets of 80%—training and 20%—testing cases, according to the stratified 10-fold cross validation sampling scheme. Each set of training cases is used to construct classification rules for each mining method.

In Table 3, using test data sets, the average accuracy results of each mining method, yielded by experimentation are presented. Such accuracy results are computed by 1.0 minus mean of square error (MSE). All the mining methods generate over 95% accuracy rates on the Wisconsin breast cancer (WBC) data set. This is due to the simple charac-

Table 2 Characteristics of considered medical data sets

Dataset	#Cases	#Attri.	#Real	#Nominal	#Class	Dev. Cla. (%)	Maj. Cla. (%)	Min. Cla. (%)
WBC	699	9	–	9	2	21.95	65.52	34.48
ECHO	131	11	8	3	2	24.29	67.18	32.82
PIMA	768	8	2	6	2	21.36	65.10	34.90
PKS	132	22	22	–	2	35.90	69.77	13.95
TG	215	5	4	1	3	39.47	69.77	13.95
LYMPH	148	18	–	18	4	37.74	54.73	1.35

Table 3 The average accuracy yielded by the five mining methods

Dataset	Method	Average accuracy
WBC	OMFAM	**96.86**
	MFAM	95.57
	FAM	93.71
	SVM	96.21
	ANFIS	94.57
ECHO	OMFAM	**93.85**
	MFAM	87.31
	FAM	82.31
	SVM	84.62
	ANFIS	79.23
PIMA	OMFAM	**75.78**
	MFAM	72.14
	FAM	72.86
	SVM	63.05
	ANFIS	75.4
PKS	OMFAM	**89.49**
	MFAM	83.33
	FAM	82.05
	SVM	70.77
	ANFIS	87.18
TG	OMFAM	**93.49**
	MFAM	88.37
	FAM	56.05
	SVM	93.26
	ANFIS	90.23
LYMPH	OMFAM	**81.33**
	MFAM	72.00
	FAM	70.33
	SVM	80.33
	ANFIS	68.00

teristics of WBC, having only nominal attribute terms, 2 classes and low class distribution deviation as well. Contrarily, all those methods show the low efficiency on Pima Indians (PIMA) data set. This may be caused by the difficulty of diagnosis of diabetes. Diabetes patients show many symptoms and some of these symptoms appear in the other types of diseases, i.e. many diseases share symptoms. According to the Lymphography (LYMPH) data set, the best accuracy of only 81.33% is produced by OMFAM; whilst the near best, 80.33% is resulted by SVM. In accordance with LYMPH, one would see the high deviation of class distribution among four classes; consistently, the high ratio of approximately 54.73 : 1.35 which is equivalent to 41 : 1 is indicated for the majority and minority class sample cases. OMFAM obtains the highest accurate rate (89.49%) on Parkinson (PKS) data set, although PKS has similar difficulties to LYMPH. However, the higher number of classes, the higher difference between majority and minority class sample cases are distinct between those two data sets. By the criteria based on average accuracy, the superiority of OMFAM over MFAM and the others is evidence of the superiority of SA optimization over the modified fuzzy ant-miner. In addition, for most data sets, MFAM denotes better quality than FAM. This emphasizes the benefits of the modification of FAM by means of applying weights of cases and terms of attribute. According to the results shown in Table 3, SVM seems to generate competitive results with OMFAM regarding Thyroid Gland (TG) and LYMPH. Even so, there would be further performance assessment criteria that needs to be taken into account. Figure 5 denotes the best accuracy rate, yielded by OMFAM with a few deviations from median values. The figure remarks the higher accuracy of OMFAM over all the related methods. Besides, it is seen the superiority of OMFAM over MFAM for all data sets. Such a predomination of OMFAM, once again denotes the benefit of SA optimization model. Figure 6 illustrates the estimation of the overfitting property for each mining methods. The overfitting of each method is evaluated by comparing the classification precision of training and testing cases. A pair of black and white consecutively represents the accuracy of those training and testing. The figure points out the large overfitting of SVM in many circumstances. That means efficiency of SVM is mainly determined based on the training data set. In contrast with SVM, OMFAM yields very

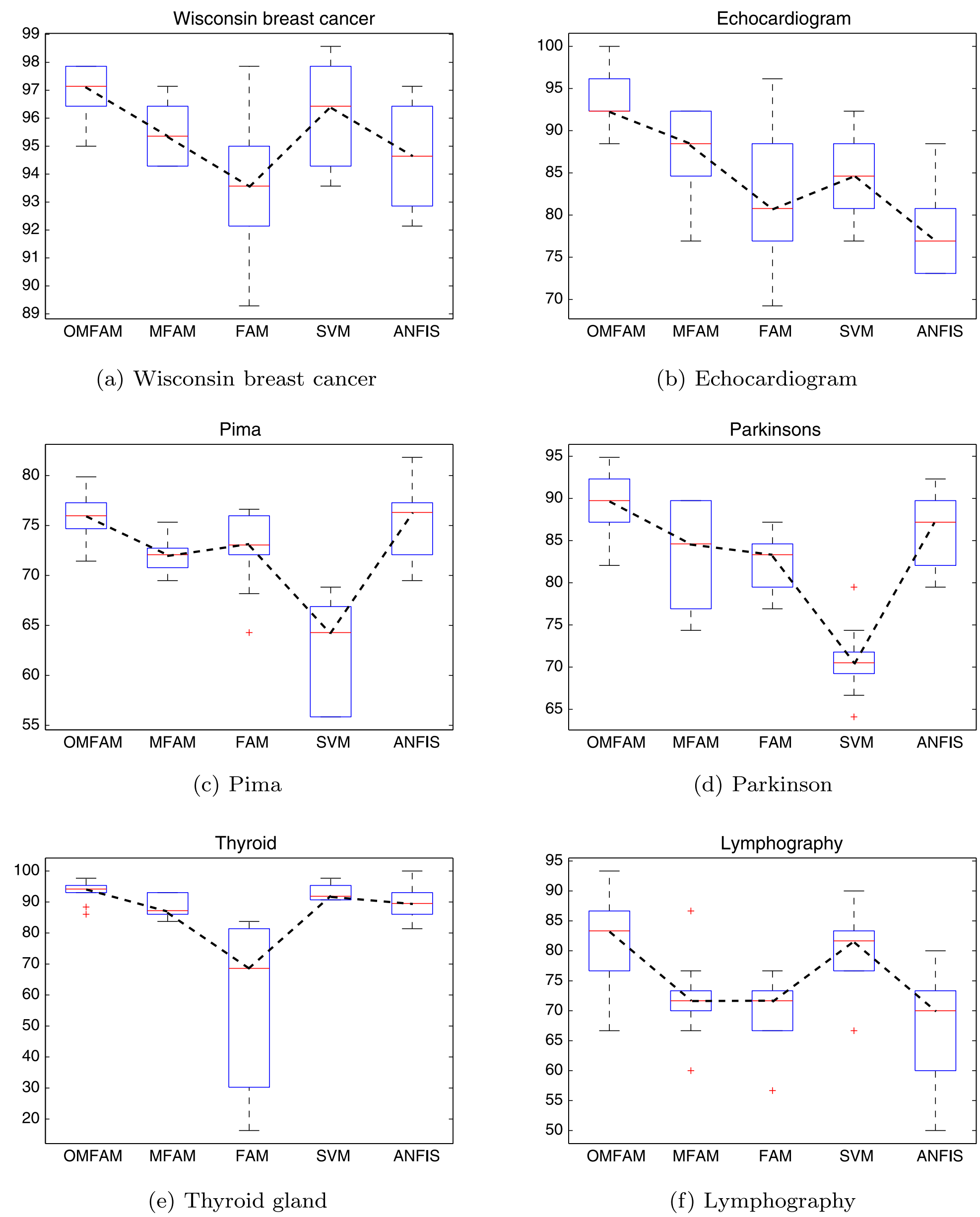
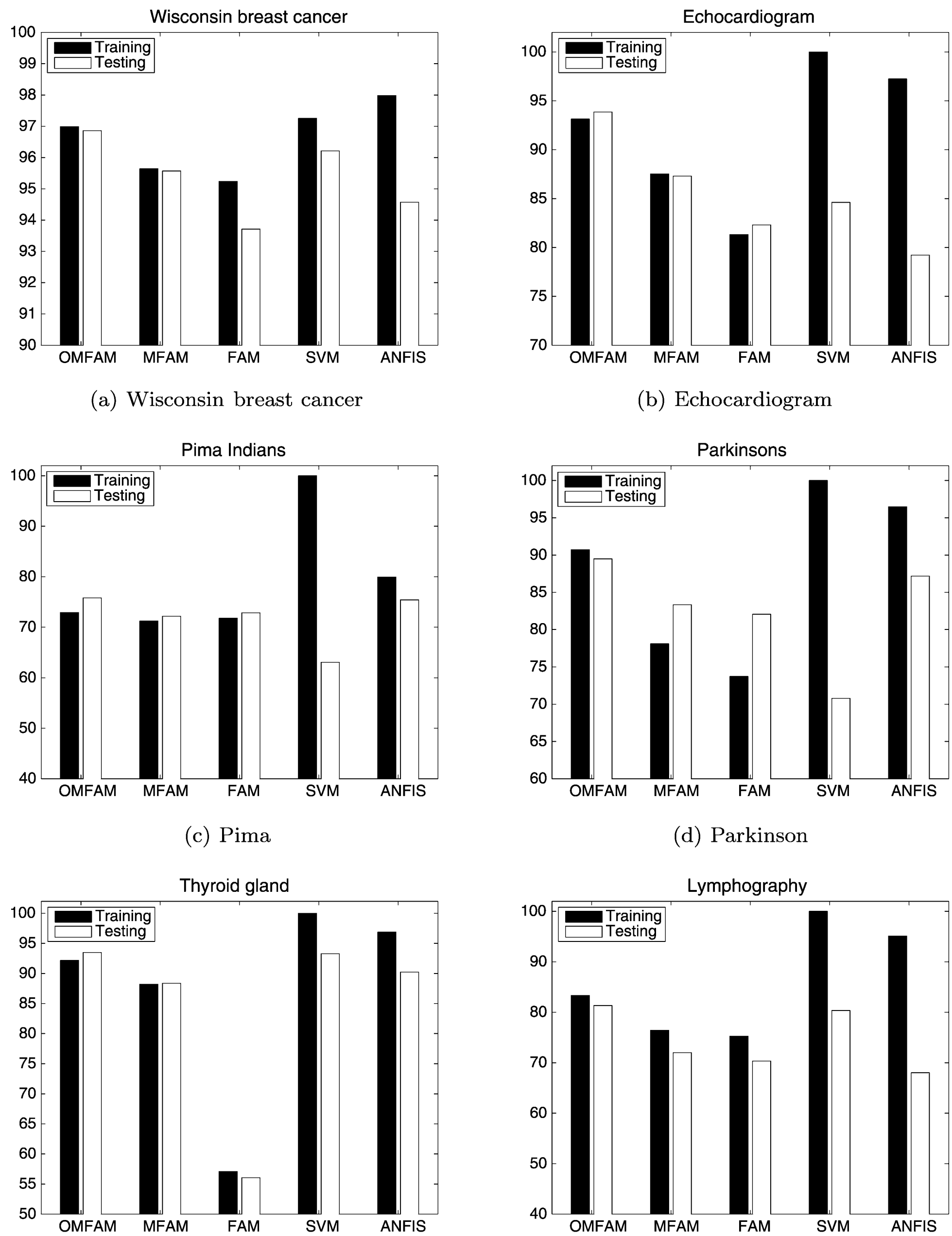


Fig. 5 Range of rule accuracy for different mining algorithms on six medical data

close training and testing results. According to most relevant cases, OMFAM remarkably shows the least overfitting.

In addition, Table 4 reports sensitivity and specificity of the mining methods, where the definition of those measurements has been described in Sect. 2. One could say sensitivity and specificity consecutively reflect the accuracy of positive and negative classes. Another precision measurement, area under the receiver operating characteristics curve (AUC) represents how well the method separates the tested data set into positive and negative classes. According to AUC, an area of 1.0 represents a perfect test; whilst 0.5 represents a worthless one. Table 4 shows SVM, having the prominent correctness of positive class identification; in contrast, having inconspicuous and low correctness of negative one. Based on AUC measure, OMFAM performed best on WBC, ECHO and PIMA, of which class distribution deviation are not extensive. It is noticeable that TG has the highest degree of class distribution deviation as well as the amount of difference between the majority class and minority class sample cases. However, OMFAM and ANFIS yield the same high level of 0.91 AUC with respect to TG. Results on PKS data set show the AUC rate of OMFAM stands on second rank after ANFIS. This is due to about 9% higher accuracy of positive class, produced by ANFIS than OMFAM; while around 3% lower accuracy of negative class than OMFAM. LYMPH has the high deviation of class distribution among four classes; and has high level of distinction between the majority and minority class sample cases. Although it indicates the best AUC rate yielded by SVM, the rival specificity rate is denoted on SVM and OMFAM.

Fig. 6 Comparison of rules accuracy based on training and the corresponding testing



(a) Wisconsin breast cancer (b) Echocardiogram

(c) Pima (d) Parkinson

(e) Thyroid gland (f) Lymphography

However, the whole results regarding Lymph data, the low degree of sensitivity along with AUC is produced by all the miners. The overall results regarding specificity and AUC results, OMFAM, FAM and ANFIS are competitive; whilst most data sets show SVM obtains the best sensitivity degree.

Besides the accuracy, the other performance assessment for the mining methods refers to rules interpretability. The measurement of the rule interpretability is shown in Table 5. According to the best rule set, generated by each classification method, the number of rules and the average number of attribute terms ('$x_d is A_{dj}$', represented by $term_{dj}$) per rule are estimated for measuring the rule interpretability. The median of classification accuracy is simultaneously considered in the same table. Wilcoxon's rank sum test is utilized for comparing a pair of the mining methods one-at-a-time: OMFAM-MFAM, MFAM-FAM and OMFAM-FAM. Such comparison test reports none of significant difference among those three mining methods in terms of the average terms per rule. However, a significant difference between OMFAM and FAM in terms of the number of rules is pointed out. The reason behind such difference is relevant to the use of the reduction of the weight of training cases in OMFAM; whereas the immediate termination of the matched case is executed in FAM. The weights reduction provides better

Table 4 Sensitivity, specificity and AUC resulted by five methods

Dataset	Method	Sensitivity (%)	Specificity (%)	AUC
WBC	OMFAM	96.89	**96.58**	**0.97**
	MFAM	96.48	93.88	0.95
	FAM	97.77	85.78	0.92
	SVM	**98.10**	92.40	0.95
	ANFIS	97.35	91.21	0.94
ECHO	OMFAM	98.82	**79.73**	**0.89**
	MFAM	93.21	72.05	0.83
	FAM	85.15	70.80	0.78
	SVM	**100.00**	60.00	0.80
	ANFIS	84.70	74.88	0.78
PIMA	OMFAM	90.55	48.51	**0.70**
	MFAM	88.77	42.67	0.66
	FAM	82.24	**56.03**	0.69
	SVM	**93.54**	8.91	0.51
	ANFIS	88.91	52.59	0.67
PKS	OMFAM	70.16	94.36	0.82
	MFAM	35	93.28	0.64
	FAM	21.25	**95.20**	0.58
	SVM	**92.44**	65.73	0.79
	ANFIS	79.07	91.25	**0.88**
TG	OMFAM	**99.01**	80.88	**0.91**
	MFAM	52.74	75.38	0.64
	FAM	51.60	72.52	0.62
	SVM	84.21	92.97	0.89
	ANFIS	84.83	**94.89**	**0.91**
LYMPH	OMFAM	45.45	**90.58**	0.67
	MFAM	42.28	87.78	0.64
	FAM	41.41	86.34	0.63
	SVM	**59.47**	90.41	**0.75**
	ANFIS	52.80	86.12	0.68

chance for OMFAM to remain more cases for constructing the rules than immediate cases termination in FAM. Although OMFAM may have some irrelevant terms resided in the best mined rules, the better median of accuracy with regard to OMFAM than that of FAM indicates the merit of SA optimization on OMFAM. In Table 5, neither resulted rules of ANFIS nor SVM are not applicable. The reason is that SVM does not yield interpretable rule-based output; instead, it generates incomprehensible support vectors as classification boundaries. Similar to ANFIS, a fixed number of rules as well as number of attribute terms per rule is pre-specified; and is never changed during the rules generation process.

In addition, the details of the mined rules yielded by OMFAM are displayed in Table 6. The elements of the best rule sets along with the percentages of the used attribute terms are revealed for each data set. Four of the six medical data

Table 5 Number of rules and average number of terms per rule

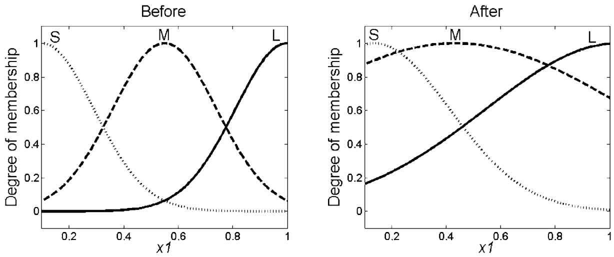
Dataset	Method	Median of classification accuracy	No. of rules/rule set	Terms/rule
WBC	OMFAM	**97.14**	3	1.67
	MFAM	95.36	**2**	2
	FAM	93.57	**2**	**1.5**
ECHO	OMFAM	**92.31**	**3**	2.67
	MFAM	88.46	**3**	1.67
	FAM	80.77	4	**1.5**
PIMA	OMFAM	**75.97**	3	2
	MFAM	72.08	3	2
	FAM	73.05	**2**	**1.5**
PKS	OMFAM	**89.74**	3	1.33
	MFAM	84.62	5	1.2
	FAM	83.33	**2**	**1**
TG	OMFAM	**94.19**	3.6	2.16
	MFAM	87.21	**3**	2.33
	FAM	68.60	**3**	**1.33**
LYMPH	OMFAM	**83.33**	4.5	2.37
	MFAM	71.67	8	**2.13**
	FAM	71.67	**2**	2.5

sets, WBC, ECHO, PKS and LYMPH show that the lower than half of the number of the attribute terms in the rule set is in used. It has been demonstrated that a few number of data set is obtained by rule pruning. Such rule pruning is gained by the influence of rule quality outcome. The successfulness of the OMFAM in terms of high quality of the mined rules mostly depends on SA optimization.

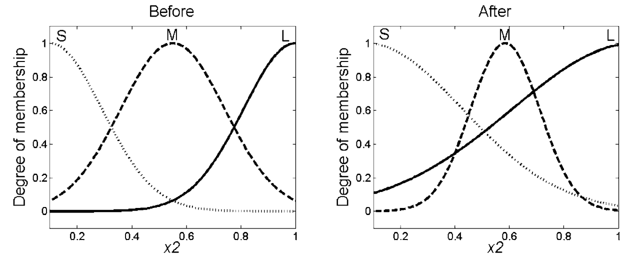
Figures 7, 8, 9–12 express diagrams, representing the changes in fuzzy set functions of all used attributes before and after the SA optimization. Such changes are the consequences of the adaptation of fuzzy set parameters (center and standard deviations), caused by SA optimization. The diagrams are employed for perceiving the effect of SA optimization on fuzzy parameters within OMFAM compared to the non-optimized approach MFAM. Such SA optimization on the fuzzy set parameters can be determined against the fixed fuzzy sets consecutively in left and right diagrams contained in each figure. One could find in Table 3 the percentages of the increase of accuracy rate made by SA optimization on OMFAM. The relationship analysis should be considered between the diagrams in Figs. 7–9, 10, 11, 12 and such increase percentages of accuracy rate of OMFAM over MFAM in Table 3. In WBC data set, the SA optimization causes the variation of fuzzy set functions, appearing in all attributes in the rule set. Even so, a few differences of accuracy rate between OMFAM and MFAM is reported, it is obvious that most of classification methods can obtain near-optimal solutions for WBC. Approximately,

Table 6 The best rule sets as well as attributes and their percentage used in regard to OMFAM

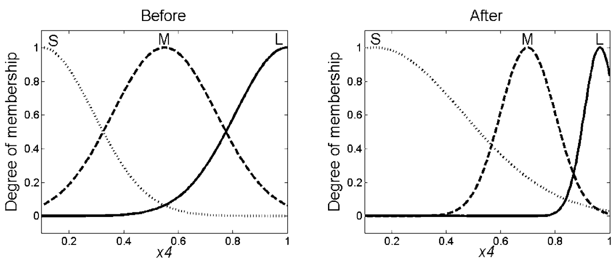
Dataset	Best rules	Attributes used in the best rules	Percentages of attributes used (%)
WBC	IF x_2 = 'S' and x_4 = 'S' and x_6 = 'S' THEN class = 0 IF x_6 = 'L' THEN class = 1 IF x_1 = 'L' THEN class = 1	x_1: Clump Thickness x_2: Uniformity of Cell Size x_4: Marginal Adhesion x_6: Bare Nuclei	33.33
ECHO	IF x_3 = 'S' and x_8 = 'M' and x_9 = 'M' and x_{11} = 'S' THEN class = 0 IF x_8 = 'M' and x_9 = 'M' and x_{11} = 'S' THEN class = 0 IF x_1 = 'S' THEN class = 1	x_1: survival x_3: age-at-heart-attack x_8: wall-motion-score x_9: wall-motion-index x_{11}: group	45.45
PIMA	IF x_2 = 'M' and x_5 = 'S' and x_7 = 'S' THEN class = 0 IF x_1 = 'S' and x_3 = 'M' and x_5 = 'S' THEN class = 0 IF x_7 = 'S' and x_8 = 'S' THEN class = 0 IF x_8 = 'M' THEN class = 1 IF x_2 = 'L' THEN class = 1	x_1: Number of times pregnant x_2: Plasma glucose concentration a 2 hours in an oral glucose tolerance test x_3: Diastolic blood pressure (mm Hg) x_5: 2-Hour serum insulin (mu U/ml) x_7: Diabetes pedigree function x_8: Age (years)	75.00
PKS	IF x_{20} = 'S' THEN class = 0 IF x_1 = 'S' THEN class = 1 IF x_2 = 'S' and x_{16} = 'M' THEN class = 1	x_1: MDVP:Fo(Hz) x_2: MDVP:Fhi(Hz) x_{16}: HNR x_{20}: spread1	18.18
TG	IF x_3 = 'S' and x_4 = 'S' and x_5 = 'S' THEN class = 0 IF x_2 = 'L' THEN class = 1 IF x_1 = 'S' and x_4 = 'S' and x_5 = 'S' THEN class = 1 IF x_2 = 'S' THEN class = 2	x_1: T3-resin uptake test x_2: Total serum thyroxin x_3: Total serum triiodothyronine x_4: basal thyroid-stimulating hormone (TSH) x_5: Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropin-releasing hormone	100.00
LYMPH	IF x_{18} = 'S' THEN class = 1 IF x_{13} = 'M' THEN class = 1 IF x_8 = 'L' and x_9 = 'S' and x_{17} = 'M' THEN class = 2 IF x_8 = 'M' and x_{15} = 'L' and x_{18} = 'L' THEN class = 3 IF x_{11} = 'S' THEN class = 3	x_8: early uptake x_9: lym.nodes dimin x_{11}: changes in lym x_{13}: changes in node x_{15}: special forms x_{17}: dislocation x_{18}: no. of nodes	33.33



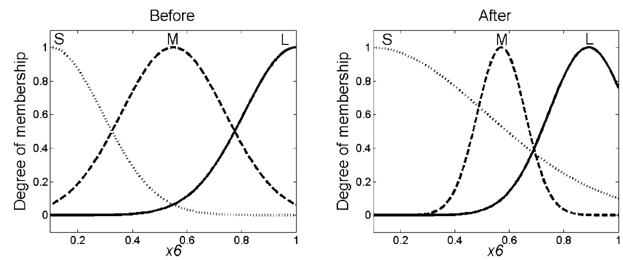
(a) x_1: Clump Thickness



(b) x_2: Uniformity of Cell Size

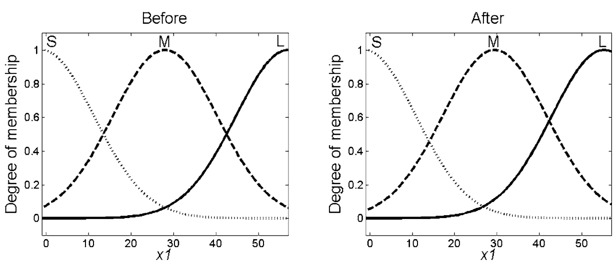


(c) x_4: Marginal Adhesion

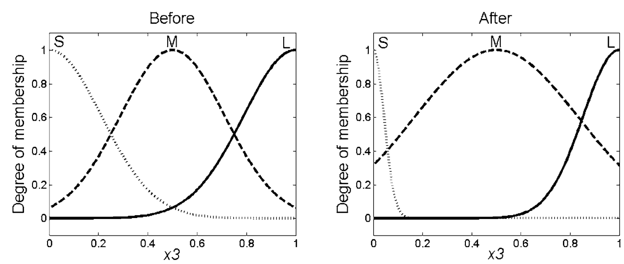


(d) x_6: Bare Nuclei

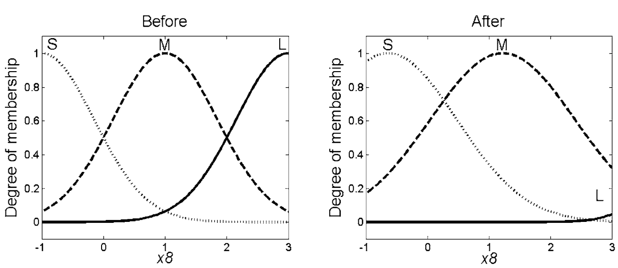
Fig. 7 The effect of SA optimization on fuzzy parameters in OMFAM, using Wisconsin breast cancer data set



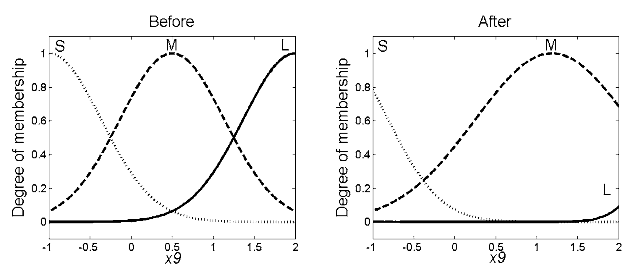
(a) x_1: survival



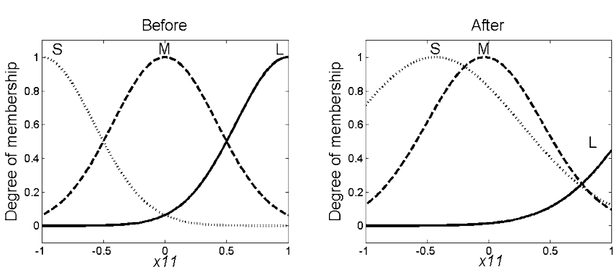
(b) x_3: age-at-heart-attack



(c) x_8: wall-motion-score

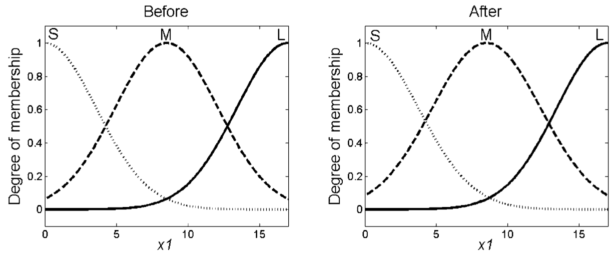


(d) x_9: wall-motion-index

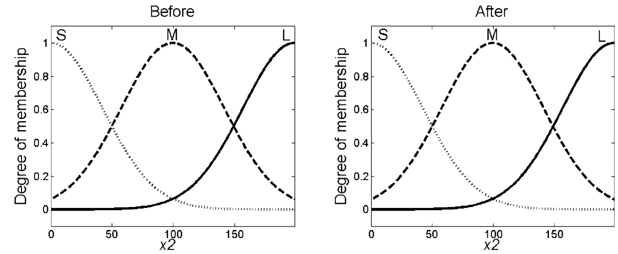


(e) x_11: group

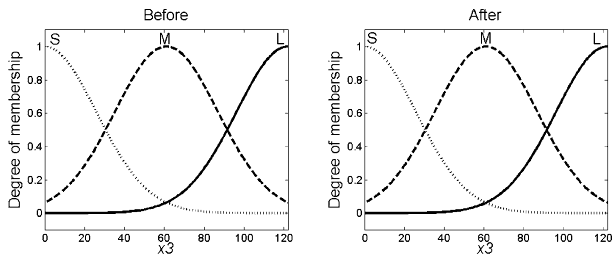
Fig. 8 The effect of SA optimization on fuzzy parameters in OMFAM, using echocardiogram data set



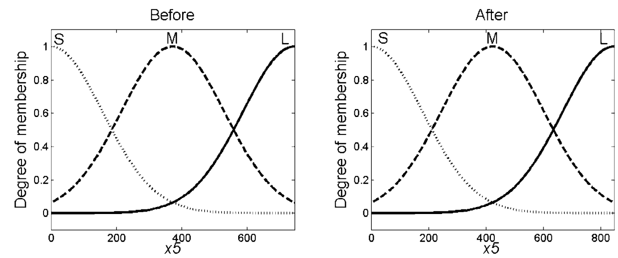
(a) x_1: Number of times pregnant



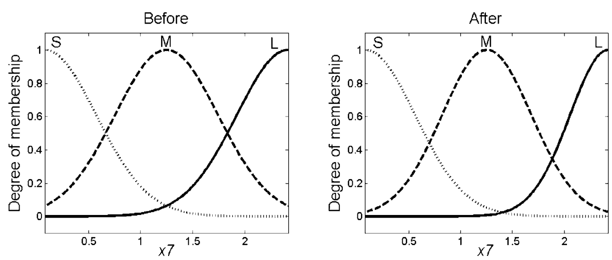
(b) x_2: Plasma glucose concentration a 2 hours in an oral glucose tolerance test



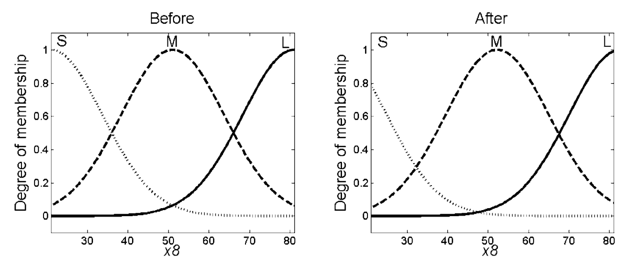
(c) x_3: Diastolic blood pressure (mm Hg)



(d) x_5: 2-Hour serum insulin (mu U/ml)

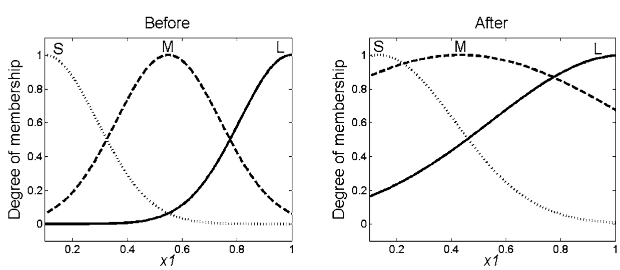


(e) x_7: Diabetes pedigree function

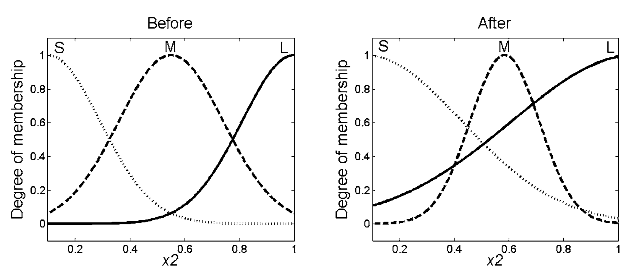


(f) x_8: Age (years)

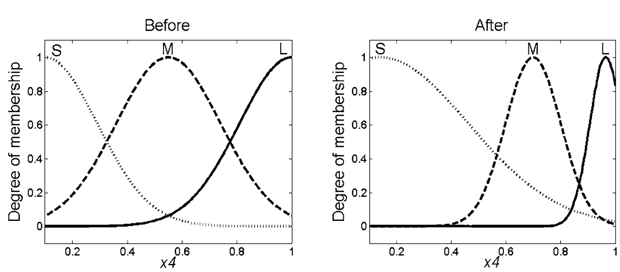
Fig. 9 The effect of SA optimization on fuzzy parameters in OMFAM, using Pima Indians diabetes data set



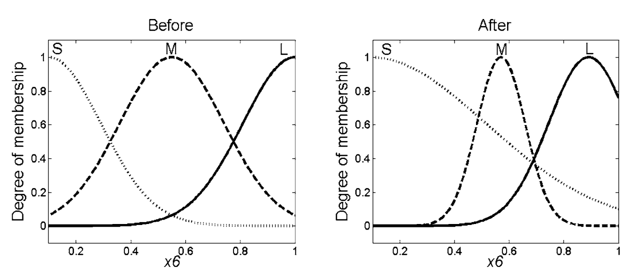
(a) x_1: MDVP:Fo(Hz)



(b) x_2: MDVP:Fhi(Hz)

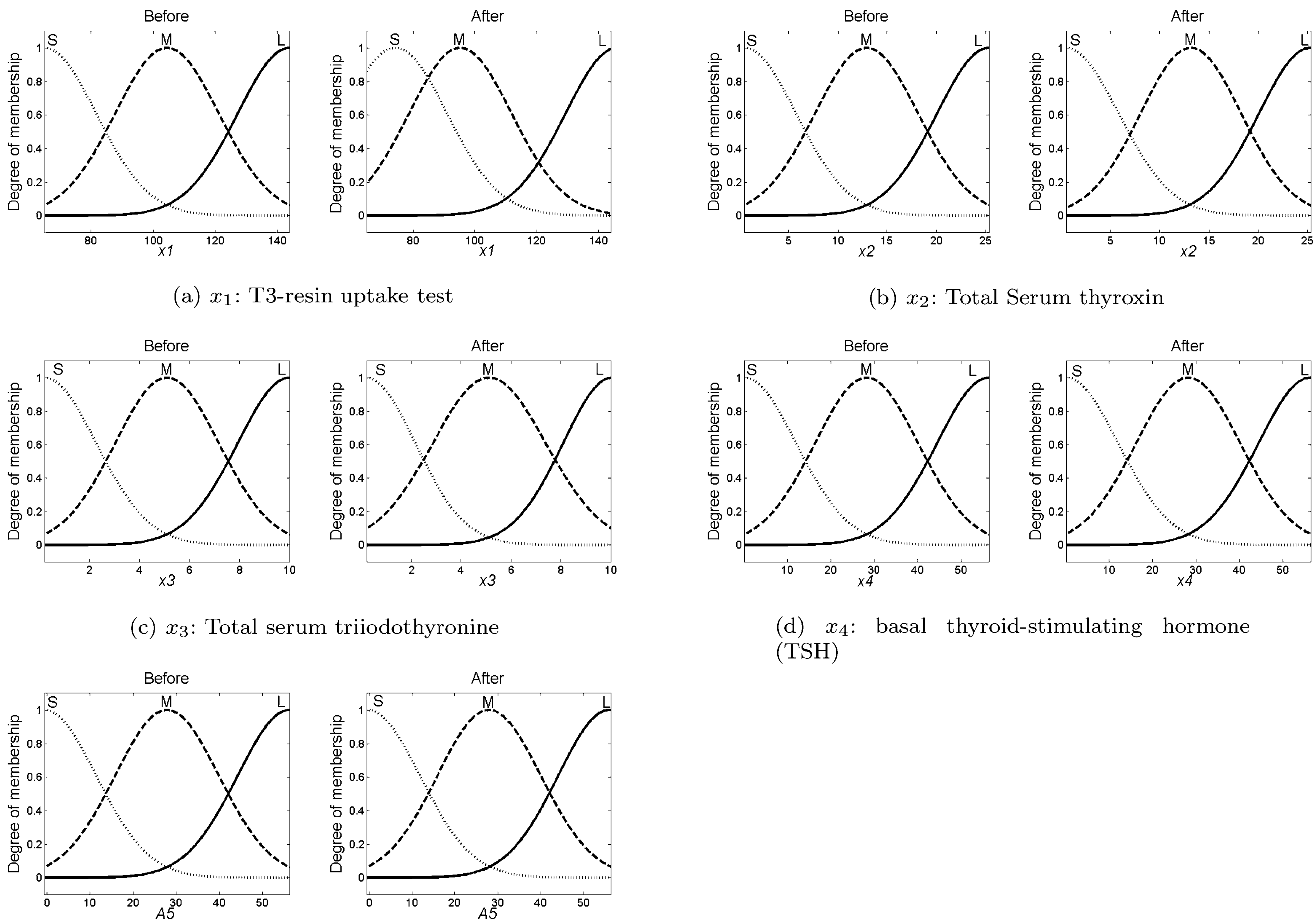


(c) x_3: HNR



(d) x_6: spread1

Fig. 10 The effect of SA optimization on fuzzy parameters in OMFAM, using Parkinson data set



(a) x_1: T3-resin uptake test

(b) x_2: Total Serum thyroxin

(c) x_3: Total serum triiodothyronine

(d) x_4: basal thyroid-stimulating hormone (TSH)

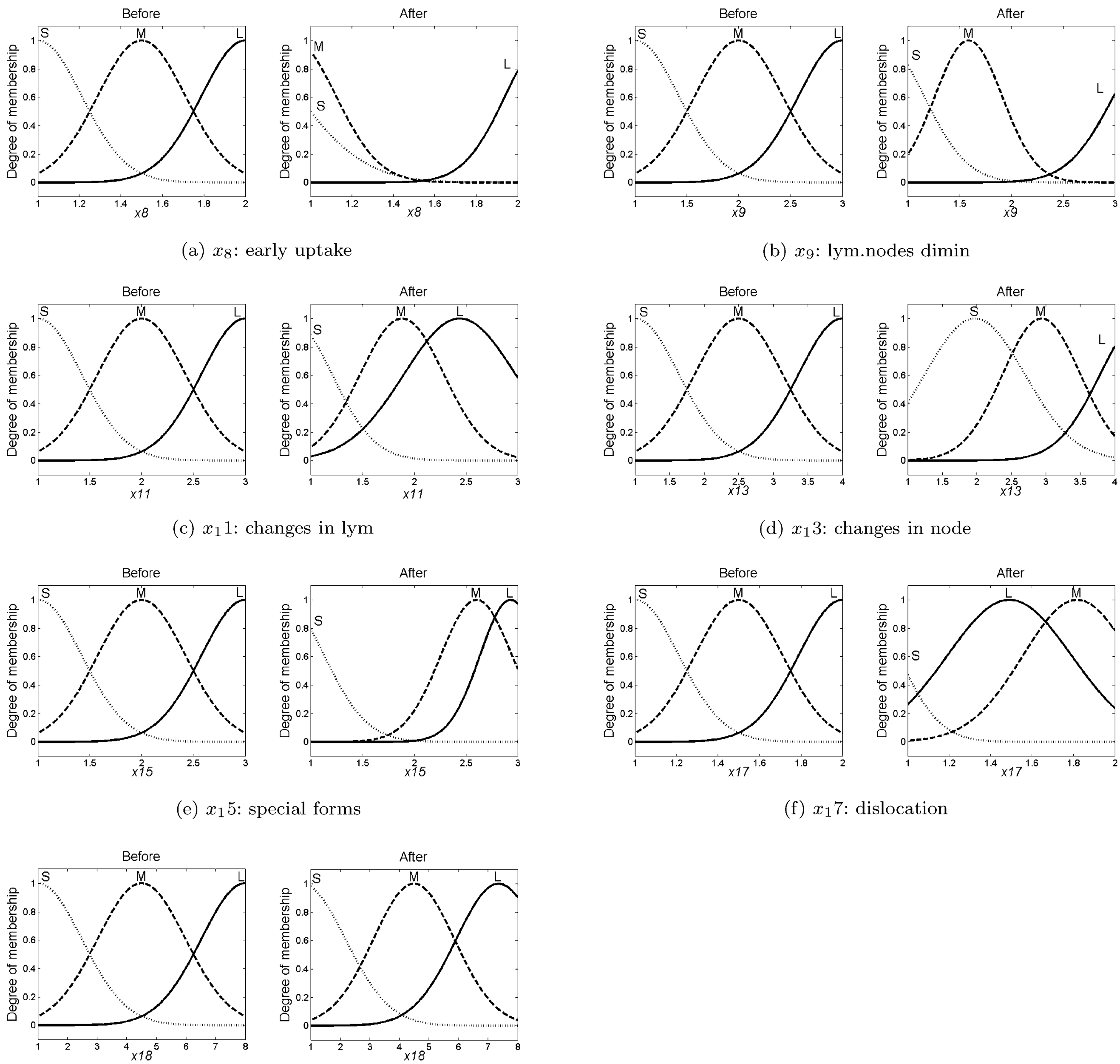
(e) x_5: Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropin-releasing hormone

Fig. 11 The effect of SA optimization on fuzzy parameters in OMFAM, using Thyroid data set

7.49% accuracy improvement is produced by using OMFAM, instead of MFAM with respect to ECHO data set. Four out of five fuzzy sets alter from those of originals as seen in the diagram. Only x1: survival that seems to be unchanged. This is one of the evidences of the advantage belonging to SA optimization. For PIMA data set, even if a few amount of fuzzy set's modification, caused by the optimization is appears, about 5.05% better quality than MFAM is produced by OMFAM. In TG, the effect of the optimization lies on only x1: T3-resin uptake test; even so, 5.79% superior quality of OMFAM than MFAM is reported. A certain degree of alteration is pointed out for all attributes in PKS and almost all in LYMPH; about 7.39% and 12.96% of higher quality, sequentially yielded by OMFAM than those of MFAM emphasize the goodness of SA optimization.

6 Discussion

The computational study, conducted in this work relies on two main performance measurements; accurateness and interpretability. The results, in terms of accurateness indicate the credibility of OMFAM that uses the merit of SA for optimizing modified fuzzy ant miner. Such credibility is pointed out by the higher quality of OMFAM than MFAM and FAM as well as the efficient neural classifiers i.e., SVM and ANFIS. This section concerns with the comparison of OMFAM against other methods in the literature that earlier mentioned in the related work section. For Wisconsin breast cancer, Fc-AntMiner [32] and AntMiner-C [33] reported the better results than that of OMFAM. However, Fc-AntMiner employed 30 number of ants and 6 fuzzy sets for each attribute term achieved 97.51% accuracy, whereas AntMiner-



(a) x_8: early uptake (b) x_9: lym.nodes dimin

(c) $x_1$1: changes in lym (d) $x_1$3: changes in node

(e) $x_1$5: special forms (f) $x_1$7: dislocation

(g) $x_1$8: no. of nodes

Fig. 12 The effect of SA optimization on fuzzy parameters in OMFAM, using Lymphography data set

C used 1,000 ants obtained 97.54%; whilst OMFAM uses only 5 ants and 3 fuzzy sets. For the mining method, the parameter setting is considered as the major cause of the outcome. By this reason, the additional experiment, using 30 number of ants but only 3 fuzzy sets in OMFAM is executed for the same data set. The result shows 98.39% degree of accuracy which is higher than or comparable to those of the two literatures. Similarly, as per Thyroid gland data set, the powerful classifiers in the literature e.g., SIM [41], ESTDD [39] and PNN [38] sequentially reached 96.86%, 95.33% and 94.81% of accuracy degree. OMFAM with only 5 ants obtains 93.44%; however, using 30 ants, it achieves 96.92%. Therefore, OMFAM indicates better quality than or comparable performance to those three effective classification methods. Additionally, three efficient methods in literature i.e., TACO-miner [42], adaptive ANN using artificial immune system (ANN-AIS) [43] and also SVM incorporated with feature discrimination (FDSVM) [44] are considered here for Echocardiogram data set. The outcome, using Echocardiogram data set is like the one, using Thyroid gland. TACO-miner, using 100 number of ants presented 96.4% degree of quality; ANN-AIS, executed within

5,000 maximum iterations reported 94.59%; whilst FDSVM yielded 87.69% accuracy. Whereas OMFAM, using 30 ants produces 96.92%.

The results of OMFAM, stated above manifest the importance of assigning different number of ants (5 or 30 ants). It is notable that fuzzy ant-mining approaches in literature, even those recent ones have employed a large number of ants to mine the rules. OMFAM, utilizing SA for optimizing the modified fuzzy ant-miner shows the effective results even if the small number of 30 ants is used for extracting the rules. It is rather unnecessary to directly compare OMFAM method with other existing classification methods. This is because the works in the literature, just-earlier compared to OMFAM have reported an extensive comparison study against other proficient classification methods. Fc-AntMiner as well as AntMiner-C, executed for Wisconsin breast cancer data set declared better classification quality than several successful classifiers e.g., SVM, C4.5, Naive Bayes, Neural network, KNN and Bayesian network as well as ANTMINER, RIPPER and Logistic regression. SIM, ESTDD and PNN classification methods were employed for Thyroid data set. The quality of their results outperformed the competent mining methods e.g., Multi-layer perceptron (MLP) with back propagation, RBF neural network, Learning vector quantization (LVQ), linear discriminant analysis (LDA) and C4.5. Furthermore, TACO-miner and ANN-AIS illustrated better accuracy against the C5.0 and GANN-C [42]. The latter method combines chaotic dynamics, genetic algorithm, and neural network. Thus, by indirect induction, it is reasonable to conclude that the OMFAM will compare favorably with several existing fuzzy ant classifiers.

7 Conclusion

The objective of this paper is to illustrate the ability of simulates annealing (SA) to develop an accurate modified fuzzy ant classifier. Employing SA for classification tasks gains effective exploration and exploitation in the large search space. SA optimization is also used to find the global optimum set of fuzzy if-then rules, usually utilized in medical applications. Experiments are performed with six UCI medical data sets. The comparison tests are executed on fuzzy ant-mining approaches, modified fuzzy ant-miner (MFAM) and traditional fuzzy ant-miner (FAM) as well as efficient neural networks, SVM and ANFIS. The performance measurements concern with accurateness and interpretability of the resulted mined rules. Contribution in terms of accurateness of OMFAM, optimized by SA is confirmed by such evidences as the best average and median of accuracy with low standard deviation, the highest AUC for five out of six data sets and the least overfitting of classification as well. On the other side, OMFAM, MFAM and FAM yield comparable results with regard to the degree of interpretability, concerning number of resulted rules. Although the high quality, regarding the terms per rules is indicated by FAM, the results indicate that the proposed OMFAM with SA optimization achieves competitive results in comparison with several well-known classification algorithms. Nonetheless, OMFAM consumes high training time. This problem could be addressed according to intelligently adjusting cooling rate or decrease rate of temperature in SA. Even so, if the training process leads to a better classification, it is still practically important since the training is only required to be performed once.

In the future, some other factors which affect the accurateness and interpretability of the mined rules can be considered in OMFAM. Moreover, other advanced searching technique for determining suitable fuzzy set parameters can be used to improve classification accuracy.

Acknowledgements We gratefully acknowledge the financial support from Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (PHD/0221/2550) to Thannob Aribarg, Siriporn Supratid and Chidchanok Lursinsap.

References

1. Michalski RS, Bratko I, Kubat M (1998) Machine learning and data mining: methods and applications. Wiley, New York
2. Steimann F (2001) On the use and usefulness of fuzzy sets in medical AI. Artif Intell Med 21:131–137
3. Leung KS, Felix Wong WS, Lam W (1989) Applications of a novel fuzzy expert system shell. Expert Syst 6:2–10. doi:10.1111/j.1468-0394.1989.tb00070.x
4. Liao SH (2005) Expert systems methodologies and applications—a decade review form 1995 to 2004. Expert Syst Appl 28:93–103. doi:10.1016/j.eswa.2004.08.003
5. Ilias M, Elias Z, Ioannis A (2009) An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Appl Intell 30(1):24–36. doi:10.1007/s10489-007-0073-z
6. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
7. Huang S-J, Chiu N-H (2009) Applying fuzzy neural network to estimate software development effort. Appl Intell 30:73–83. doi:10.1007/s10489-007-0097-4
8. Ishibuchi H, Nakashima T, Murata T (1999) Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Trans Syst Man Cybern, Part B, Cybern 29(5):601–618
9. Nozaki K, Ishibuchi H, Tanaka H (1996) Adaptive fuzzy rule-base classification systems. IEEE Trans Fuzzy Syst 4(3):238–250
10. Shi Y, Eberhart R, Chen Y (1989) Implementation of evolutionary fuzzy systems. IEEE Trans Fuzzy Syst 7(2):109–119
11. Young M (2002) The technical writers handbook. University Science, Mill Valley
12. Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665–685. doi:10.1109/21.256541
13. Chang BR, Tsai H-F (2009) Quantum minimization for adapting ANFIS outputs to its nonlinear generalized autoregressive conditional heteroscedasticity. Appl Intell 31(1):31–46. doi:10.1007/s10489-007-0110-y

14. Ubeyli ED (2009) Adaptive neuro-fuzzy inference systems for automatic detection of breast cancer. J Med Syst 33:353–358. doi:10.1007/s10916-008-9197-x
15. Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets Syst 141:59–88
16. Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H (1995) Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans Fuzzy Syst 3(3):260–271
17. Roubos H, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction. IEEE Trans Fuzzy Syst 9(4):516–524
18. Mohamadi H, Habibi J, Abadeh MS, Saadi H (2008) Data mining with a simulated annealing based fuzzy classification system. Pattern Recognit 41:1824–1833
19. Saniee AM, Habibi J, Soroush E (2008) Induction of fuzzy classification systems via evolutionary ACO-based algorithms. Int J Simul Syst Sci Technol 9(3):1–8
20. Saniee AM, Habibi J, Lucas C (2007) Intrusion detection using a fuzzy genetics-based learning algorithm. J Netw Comput Appl 30:414–428
21. Saniee AM, Habibi J, Soroush E (2007) Induction of fuzzy classification systems using evolutionary ACO-based algorithms. In: Proceedings of the first Asia international conference on modelling and simulation (AMS'07). IEEE Press, New York
22. Dorigo M, Stutzle T (2004) Ant colony optimization. MIT Press, Cambridge
23. Dorigo M, Blum C (2005) Ant colony optimization theory: a survey. Theor Comput Sci 344:243–278
24. Dorigo M, Maniezzo V, Colorni A (1996) The ant system: Optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern 26:1–13
25. Jessica R, Dolores C, Javier C, Pedro I (2011) Using the ACO algorithm for path searches in social networks. Appl Intell. doi:10.1007/s10489-011-0304-1
26. Blum C (2005) Review ant colony optimization: introduction and recent trends. Phys Life Rev 2:353–373
27. Parpinelli RS, Lopes HS, Freitas AA (2002) Data mining with an ant colony optimization algorithm. IEEE Trans Evol Comput 6:321–332
28. Liu B, Abbass HA, McKay B (2002) Density-based heuristic for rule discovery with ant-miner. In: The 6th Australia-Japan joint workshop on intelligent
29. Liu B, Abbass HA, McKay B (2003) Classification rule discovery with ant colony optimization. In: Proc IEEE/WIC int conf on intell agent techno
30. Martens D, De Backer M, Haesen R, Vanthienen J, Snoeck M, Baesens B (2007) Classification with ant colony optimization. IEEE Trans Evol Comput 11:651–656
31. Galea M, Shen Q (2006) Simultaneous ant colony optimization algorithms for learning linguistic fuzzy rules. In: Agraham A, Grosan C, Ramos V (eds) Swarm intelligence in data mining. Springer, Berlin, pp 75–99
32. Mostafa FG, Mohamad SA (2010) Rule based classification system for medical data mining using fuzzy ant colony optimization. In: Proceedings of the world congress on engineering and computer science (WCECS 2010), vol 1, San Francisco, USA
33. Abdul RB, Waseem S (2010) A correlation-based ant miner for classification rule discovery. Neural Comput Appl. doi:10.1007/s00521-010-0490-5
34. Alatas B, Akin E (2005) FCACO: fuzzy classification rules mining algorithm with ant colony optimization. In: ICNC, vol 3, pp 787–797
35. Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680
36. The uc irvine machine learning repository (2010) http://archive.ics.uci.edu/ml/. Accessed 8 June 2010
37. Ghazavi SN, Liao TW (2008) Medical data mining by fuzzy modeling with selected features. Artif Intell Med 43(3):195–206
38. Feyzullah T (2009) A comparative study on thyroid disease diagnosis using neural networks. Expert Syst Appl 36(1):944–949
39. Ali K, Ayturk K (2008) ESTDD: expert system for thyroid diseases diagnosis. Expert Syst Appl 34(1):242–246
40. Esin D, Akif D, Derya A (2011) An expert system based on generalized discriminant analysis and wavelet support vector machine for diagnosis of thyroid diseases. Expert Syst Appl 38(1):146–150
41. Luukka P, Leppalampi T (2006) Similarity classifier with generalized mean applied to medical data. Comput Biol Med 36(9):1026–1040
42. Ozbakir L, Baykasoglu A, Kulluk S (2008) Rule extraction from neural networks via ant colony algorithm for data mining applications. In: Maniezzo V et al (eds) Proceedings of the 2nd international conference on learning and intelligent optimization-LION 2007. Lecture notes in computer science, vol 5313. Springer, Berlin, pp 177–191
43. Kahramanli H, Allahverdi N (2009) Rule extraction from trained adaptive neural networks using artificial immune systems. Expert Syst Appl 36:1513–1522
44. Yunyun W, Songcan C, Hui X (2011) Support vector machine incorporated with feature discrimination. Expert Syst Appl 38(10):12506–12513
45. Bach A (1990) Boltzmann's probability distribution of 1877. Arch Hist Exact Sci 41:1–40. doi:10.1007/BF00348700
46. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computating machines. J Chem Phys 21:1087–1091. doi:10.1063/1.1699114
47. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148(3):839–843