

RESEARCH

Open Access

Extraction methods for uncertain inference rules by ant colony optimization

Ling Chen, Yun Sun* and Yuanguo Zhu

*Correspondence:
chinalsy_881220@163.com
School of Science, Nanjing
University of Science and
Technology, Nanjing 210094, China

Abstract

In recent years, the research on data mining methods has received increasing attention. In this paper, we design an uncertain system with the extracted uncertain inference rules to solve the classification problems in data mining. And then, two extraction methods integrated with ant colony optimization are proposed for the generation of the uncertain inference rules. Finally, two applications are given to verify the effectiveness and superiority of the proposed methods.

Keywords: Uncertain inference rule; Uncertain system; Ant colony optimization algorithm; Rules extraction; Data classification

Introduction

Nowadays, databases and computer networks, coupled with the use of advanced automated data generation and collection tools, are widely used in many different fields such as finance, E-commerce, logistics, etc. As a result, the amount of data that people have to deal with is dramatically increasing. People hope to carry out scientific research, business decision, or business management on the basis of the analysis of the existing data. However, the current data analysis tools have difficulty in processing the data in depth. To compensate for this deficiency, there come the data mining techniques. Data mining is the computational process of discovering some interesting, potentially useful patterns in large data sets. Those patterns can be concepts, rules, laws, and modes. The overall goal of data mining is to extract information from a data set and transform it into an understandable structure for further use. Data mining helps us to discover valuable information and knowledge. Data mining is applied to many fields in reality. There are many successful examples [1] of data mining in business and science research. For instance, data mining is widely used in financial data analysis, telecommunication, retail, and biomedical research. Therefore, the study of data mining technology has an important practical significance.

The main jobs of data mining are data description, data classification, data dependency, data compartment analysis, data regression, data aggregate, and data prediction. What data classification does is to find a couple of models or functions that can accurately describe the characteristics of the data sets. Then, we can identify the categories of the previously unknown data. After obtaining the models or functions from the set of training data with data mining algorithms, we use many methods to describe the output such as classification rules (if-then), decision trees, mathematical formula, and neural network.

There are a variety of approaches in data mining. For mining objects in different fields, many different specified methods are invented. The approaches we usually used are statistical methods, machine learning methods, and modern intelligent optimization methods. The statistical methods are very effective methods from the start. In addition, many other data mining methods are invented based on the statistical methods. When dealing with classification problems, Bayesian classification and Bayesian belief network are important classification methods that based on the statistical principle. Machine learning methods are mainly used to solve the conceptual learning, pattern classification, and pattern clustering problems. The core content of machine learning is inductive learning. And there already exist a number of mature technology methods, such as decision tree method for classification problems. Decision trees method is one of the most popular classification methods. The early decision trees algorithm is ID3 method. Later, based on ID3, many algorithms such as C4.5 method [2] are proposed. Besides, there are some variants of the decision trees algorithm including incremental tree structure ID4, ID5, and expandable tree structure SLIQ for massive data set.

In recent years, intelligent optimization algorithms are widely applied into data mining. Neutral network is a simulation model for complex system with nonlinear relations. It is very suitable to deal with complex nonlinear relations in spatial data. Researchers have already proposed different network models to realize the clustering, classification, regression, and pattern recognition of the data. Furthermore, many evolution algorithms such as simulated annealing algorithm are introduced into neutral network algorithm as the optimization strategies. Genetic algorithm is a global search algorithm that simulates the biological evolution and genetic mechanism. It plays an important role in optimization and classification machine learning. Mixed algorithms of genetic algorithm and other algorithms, such as decision trees, neutral network, have been applied to the data mining technology. Ant colony optimization algorithm is a bionic optimization algorithm that simulates the behavior of the ants. Based on that, a data mining technique ant-miner [3] was invented. And Herrera [4] applied it to fuzzy rules learning. However, ant colony optimization algorithm has some weakness such as slow convergence, random initial solutions. For this reason, some improved ant colony optimization algorithms are proposed. Zhu proposed an improved ant colony optimization algorithm (ACOA) [5] and a mutation ant colony optimization algorithm (MACO) [6] to speed up the algorithms and avoid the solutions getting stuck in local optimums. Hybrid genetic ant colony optimization [7] and hybrid particle swarm ant colony optimization algorithm [8] significantly improve the performance of the original ant colony optimization algorithm.

The real world is so complex that human being may face different types of indeterminacy everyday. To get a better understanding of the real world, many mathematical tools are created. One of them is probability theory which is used to model indeterminacy from samples. However, in many cases, no samples are available to estimate a probability distribution. In this situation, we have no choice but to invite some domain experts to evaluate the belief degree that each event may occur. We cannot use probability theory to deal with belief degree since human beings usually overweight unlikely events which makes the belief degrees deviate far from the frequency. In view of this, Liu [9] founded uncertainty theory based on normality axiom, duality axiom, subadditivity axiom, and product measure axiom. It has become a powerful mathematical tool dealing with indeterminacy.

Many researchers have done a lot of theoretical work related to uncertainty theory. In 2008, Liu [10] presented the uncertain differential equation. Later, the existence and uniqueness theorem was given [11]. And the stability of uncertain differential equation was discussed [12,13]. Also, some analysis and numerical methods for solving uncertain differential equation were proposed. With uncertain differential equation describing the evolution of the system, we may solve some practical problems. Peng and Yao [14] studied an option pricing models for stocks. Zhu [15] proposed an uncertain optimal control model in 2010.

In [16,17], Liu proposed and studied the uncertain systems based on the concepts of uncertain sets, membership functions, and uncertain inference rules. An uncertain system is a function from its inputs to outputs based on the uncertain inference rule. Usually, an uncertain system consists of five parts: inputs, rule-base, uncertain inference rules, expected value operator, and outputs. Following that, Gao et al. [18] generalized uncertain inference rules and described uncertain systems with them. Peng and Chen [19] proved that uncertain systems are universal approximator and then demonstrated that the uncertain controller is a reasonable tool. Gao [20] designed an uncertain inference controller that successfully balanced an inverted pendulum with 5×5 if-then rules. What is more important is that this uncertain inference controller has a good ability of robustness.

On the basis of uncertainty theory, we consider two extraction methods for uncertain inference rules by ant colony optimization algorithm. In the next section, we review the ant colony optimization algorithm and give some basic concepts about uncertain sets. Then, we formulate a model to extract inference rules based on data set. And then, we propose an extraction method for uncertain inference rules by ant colony optimization algorithm with a mutation operation. Finally, we combine the ant colony optimization algorithm with simulated annealing algorithm to speed up the extraction method. In the last section, we discuss two typical classification problems in data mining with our results.

Preliminary

In this section, we review the ant colony optimization algorithm. And then, we give some basic concepts on uncertainty sets.

Ant colony optimization algorithm

Ant colony optimization algorithm, initiated by Dorigo, is a heuristic optimization approach. It simulates the behavior of real ants when they forage for food which relies on the pheromone communication. In ant colony optimization algorithm, each path of artificial ants walking from the food sources to the nest is a candidate solution to the problem. When walking on the path, the ants will release pheromone which evaporates over time. And the artificial ants will lay down more pheromone on the path corresponding to the better solution. While one ant has many paths to go, it will make a choice according to the amount of the pheromone on the paths. The more pheromone there is on the path, the better the solution is. As a result, bad paths will disappear since the pheromone evaporates over time. And good paths will be reserved since ants walking on it increases the pheromone levels. Finally, one path which is used by most of the ants is left. Then, the optimal solution to the problem is obtained.

Consider the following optimization problem:

$$\begin{cases} \min f(x) \\ \text{s.t.} \\ g(x) \geq 0 \\ x \in D \end{cases} \tag{1}$$

where x is the decision variable in the domain D . And $f(x)$ is the objective function while $g(x)$ is the constraint function.

We can use ant colony optimization algorithm to obtain the optimal solution to the problem (1). The parameters in the algorithm are initial pheromone τ_0 , ant transfer probability p , number of ants M , pheromone evaporation rate ρ , and number of iterations T . The procedures are as follows.

Step 1 Randomly generate a feasible solution x_0 and set optimal solution $s = x_0$. Initialize all pheromone trails with the same pheromone level τ_0 . Set $k \leftarrow 0$.

Step 2 The artificial ant generates a walking path x in some probability p according to the pheromone trails. If $x \in D$, then go to Step 3; otherwise, repeat Step 2 until $x \in D$.

Step 3 Repeat Step 2 until for each ant and generate M feasible solutions. Let s_k be the best solution in this iteration.

Step 4 If $f(s_k) < f(s)$, then $s \leftarrow s_k$ and update the pheromone trails according to the optimal solution in the current iteration.

Step 5 If $k < T$, then $k \leftarrow k + 1$ and go to Step 2; otherwise, terminate.

Step 6 Report the optimal solution.

Uncertain set

Let Γ be a nonempty set and \mathcal{L} be σ -algebra over Γ . Each $\Lambda \in \mathcal{L}$ is called an event. For any Λ , $\mathcal{M}\{\Lambda\} \in [0, 1]$. The set function \mathcal{M} defined on \mathcal{L} is called an uncertain measure if it satisfies the following three axiom: $\mathcal{M}\{\Gamma\} = 1$; $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$ for any $\Lambda \in \mathcal{L}$; $\mathcal{M}\{\bigcup_{i=1}^{\infty} \Lambda_i\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}$ for all $\Lambda_1, \Lambda_2, \dots \in \mathcal{L}$. Then, the triplet $(\Gamma, \mathcal{L}, \mathcal{M})$ is called an uncertainty space [9]. The product uncertain measure \mathcal{M} is an uncertain measure satisfying $\mathcal{M}\{\prod_{i=1}^{\infty} \Lambda_k\} = \bigwedge_{i=1}^{\infty} \mathcal{M}_k\{\Lambda_k\}$, where Λ_k are arbitrarily chosen events from \mathcal{L}_k for $k = 1, 2, \dots$, respectively.

Definition 1. [16] *An uncertain set is a function ξ from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to a collection of sets of real numbers such that both $\{B \subset \xi\}$ and $\{\xi \subset B\}$ are events for any Borel set B .*

Example 1. Take $(\Gamma, \mathcal{L}, \mathcal{M})$ to be $\{\gamma_1, \gamma_2, \gamma_3\}$ with power set \mathcal{L} . Then, the set-valued function

$$\xi(\gamma) = \begin{cases} [1, 3], & \text{if } \gamma = \gamma_1 \\ [2, 4], & \text{if } \gamma = \gamma_2 \\ [3, 5], & \text{if } \gamma = \gamma_3 \end{cases}$$

is an uncertain set on $(\Gamma, \mathcal{L}, \mathcal{M})$.

Definition 2. [16] *The uncertain sets $\xi_1, \xi_2, \xi_3, \dots, \xi_n$ are said to be independent if for any Borel sets $B_1, B_2, B_3, \dots, B_n$, we have*

$$\mathcal{M} \left\{ \bigcap_{i=1}^n (\xi_i^* \subset B_i) \right\} = \bigwedge_{i=1}^n \mathcal{M} \{ \xi_i^* \subset B_i \}$$

and

$$\mathcal{M} \left\{ \bigcup_{i=1}^n (\xi_i^* \subset B_i) \right\} = \bigvee_{i=1}^n \mathcal{M} \{ \xi_i^* \subset B_i \}$$

where ξ_i^* are arbitrarily chosen from $\{ \xi_i, \xi_i^c \}$, $i = 1, 2, \dots, n$, respectively.

Definition 3. [21] *An uncertain set ξ is said to have a membership function μ if for any Borel set B of real numbers, we have*

$$\mathcal{M}\{B \subset \xi\} = \inf_{x \in B} \mu(x), \mathcal{M}\{\xi \subset B\} = 1 - \sup_{x \in B^c} \mu(x).$$

The above equations will be called measures inversion formulas.

Remark 1. When an uncertain set ξ does have a membership function μ , it follows from the first measure inversion formula that

$$\mu(x) = \mathcal{M}\{x \in \xi\}.$$

Example 2. An uncertain set ξ is called triangular if it has a membership function

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{x-c}{b-c}, & b \leq x \leq c \end{cases} \tag{2}$$

denoted by (a, b, c) where a, b, c are real numbers with $a < b < c$.

Definition 4. [21] *A membership function μ is said to be regular if there exists a point x_0 such that $\mu(x_0) = 1$, and $\mu(x)$ is unimodal about the mode x_0 . That is, $\mu(x)$ is increasing on $(-\infty, x_0]$ and decreasing on $[x_0, +\infty)$.*

Definition 5. [16] *Let ξ be an uncertain set. Then, the expected value of ξ is defined by*

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \geq r\} dr - \int_{-\infty}^0 \mathcal{M}\{\xi \leq r\} dr$$

provided that at least one of the two integrals is finite and

$$\mathcal{M}\{\xi \geq r\} = \frac{1}{2}(\mathcal{M}\{\xi \geq r\} + 1 - \mathcal{M}\{\xi < r\}),$$

$$\mathcal{M}\{\xi \leq r\} = \frac{1}{2}(\mathcal{M}\{\xi \leq r\} + 1 - \mathcal{M}\{\xi > r\}).$$

Theorem 1. [13] *Let ξ be an uncertain set with regular membership function μ . Then*

$$E[\xi] = x_0 + \frac{1}{2} \int_{x_0}^{+\infty} \mu(x) dx - \frac{1}{2} \int_{-\infty}^{x_0} \mu(x) dx, \tag{3}$$

where x_0 is a point such that $\mu(x_0) = 1$.

Example 3. Let ξ be a triangular uncertain set denoted by (a, b, c) . Then, according to Theorem 1, we have

$$E[\xi] = \frac{a + 2b + c}{4}.$$

In fact, it follows from Equations 2 and 3 that

$$\begin{aligned} E[\xi] &= b + \frac{1}{2} \int_b^c \frac{x - c}{b - c} dx - \frac{1}{2} \int_a^b \frac{x - a}{b - a} dx \\ &= b - \frac{1}{4}(b - c) - \frac{1}{4}(b - a) \\ &= \frac{a + 2b + c}{4}. \end{aligned}$$

Uncertain inference rule

Here, we introduce concepts of the uncertain inference and uncertain system. Inference rules are the key points of the inference systems. In fuzzy systems, CRI approach [22], Mamdani inference rules [23] and Takagi-Sugeno inference rules [24] are the most common used inference rules. Fuzzy if-then inference rules use fuzzy sets to describe the antecedents and the consequents. Unlike fuzzy inference, both antecedents and consequents in uncertain inference are characterized by uncertain sets. Uncertain inference [16] is a process of deriving consequences from human knowledge via uncertain set theory. First, we introduce the following inference rule.

Inference Rule 1. [16] *Let \mathbb{X} and \mathbb{Y} be two concepts. Assume a rule ‘if \mathbb{X} is an uncertain set ξ , then \mathbb{Y} is an uncertain set η ’. From \mathbb{X} is a constant a , we infer that \mathbb{Y} is an uncertain set*

$$\eta^* = \eta|_{a \in \xi}$$

which is the conditional uncertain set of η given $a \in \xi$. The inference rule is represented by

Rule: If \mathbb{X} is ξ , then \mathbb{Y} is η

From: \mathbb{X} is a constant a

Infer: \mathbb{Y} is $\eta^* = \eta|_{a \in \xi}$

Theorem 2. [16] *Let ξ and η be independent uncertain sets with membership functions μ and ν , respectively. If ξ^* is a constant a , then the Inference Rule 1 yields that η^* has a membership function*

$$\nu^*(y) = \begin{cases} \frac{\nu(y)}{\mu(a)}, & \text{if } \nu(y) < \frac{\mu(a)}{2} \\ \frac{\nu(y) + \mu(a) - 1}{\mu(a)}, & \text{if } \nu(y) > 1 - \frac{\mu(a)}{2} \\ 0.5, & \text{otherwise.} \end{cases}$$

Based on Inference Rule 1, Gao et al. [18] proposed the multi-input, multi-if-then-rule inference rules.

Inference Rule 2. [13] Let $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_m, \mathbb{Y}$ be concepts. Assume rules 'if \mathbb{X}_1 is ξ_{i1} and \dots and \mathbb{X}_m is ξ_{im} , then \mathbb{Y} is η_i ' for $i = 1, 2, \dots, k$. From \mathbb{X}_1 is a constant a_1 and \dots and \mathbb{X}_m is a constant a_m , we infer that

$$\eta^* = \sum_{i=1}^k \frac{c_i \cdot \eta_i |_{(a_1 \in \xi_{i1}) \cap (a_2 \in \xi_{i2}) \cap \dots \cap (a_m \in \xi_{im})}}{c_1 + c_2 + \dots + c_k}, \tag{4}$$

where the coefficients are determined by

$$c_i = \mathcal{M}\{(a_1 \in \xi_{i1}) \cap (a_2 \in \xi_{i2}) \cap \dots \cap (a_m \in \xi_{im})\}$$

for $i = 1, 2, \dots, k$. The inference rule is represented by

Rule 1: If \mathbb{X}_1 is ξ_{11} and \dots and \mathbb{X}_m is ξ_{1m} , then \mathbb{Y} is η_1

Rule 2: If \mathbb{X}_1 is ξ_{21} and \dots and \mathbb{X}_m is ξ_{2m} , then \mathbb{Y} is η_2

...

Rule k: If \mathbb{X}_1 is ξ_{k1} and \dots and \mathbb{X}_m is ξ_{km} , then \mathbb{Y} is η_k

From: \mathbb{X}_1 is a_1 and \dots and \mathbb{X}_m is a_m

Infer: \mathbb{Y} is determined by Eq. (4)

Theorem 3. [13] Assume $\xi_{i1}, \xi_{i2}, \dots, \xi_{im}, \eta_i$ are independent uncertain sets with membership functions $\mu_{i1}, \mu_{i2}, \dots, \mu_{im}, v_i, i = 1, 2, \dots, k$, respectively. If $\xi_1^*, \xi_2^*, \dots, \xi_m^*$ are constants a_1, a_2, \dots, a_m , respectively, then the Inference Rule 2 yields

$$\eta^* = \sum_{i=1}^k \frac{c_i \cdot \eta_i^*}{c_1 + c_2 + \dots + c_k}$$

where η_i^* are uncertain sets whose membership functions are given by

$$v_i^*(y) = \begin{cases} \frac{v_i(y)}{c_i}, & \text{if } v_i(y) < \frac{c_i}{2} \\ \frac{v_i(y)+c_i-1}{\mu(a)}, & \text{if } v_i(y) > 1 - \frac{c_i}{2} \\ 0.5, & \text{otherwise} \end{cases}$$

and $c_i = \min_{1 \leq l \leq m} \mu_{il}(a_l)$ are constants.

Uncertain system

Uncertain system, proposed by Liu [16], is a function from its inputs to outputs based on the uncertain inference rule. Usually, an uncertain system consists of five parts: inputs that are crisp data to be fed into the uncertain system; a rule-base that contains a set of if-then rules provided by the experts; an uncertain inference rule that infers uncertain consequents from the uncertain antecedents; an expected value operator that converts the uncertain consequents to crisp values; and outputs that are crisp data yielded from the expected value operator.

Now, we consider an uncertain system with m crisp inputs $\alpha_1, \alpha_2, \dots, \alpha_m$, and n crisp outputs $\beta_1, \beta_2, \dots, \beta_n$. We have the following if-then rules:

If \mathbb{X}_1 is ξ_{11} and \dots and \mathbb{X}_m is ξ_{1m} , then \mathbb{Y}_1 is η_{11} and \mathbb{Y}_2 is η_{12} and \dots and \mathbb{Y}_n is η_{1n}

If \mathbb{X}_1 is ξ_{21} and \dots and \mathbb{X}_m is ξ_{2m} , then \mathbb{Y}_1 is η_{21} and \mathbb{Y}_2 is η_{22} and \dots and \mathbb{Y}_n is η_{2n}

...

If \mathbb{X}_1 is ξ_{k1} and \dots and \mathbb{X}_m is ξ_{km} , then \mathbb{Y}_1 is η_{k1} and \mathbb{Y}_2 is η_{k2} and \dots and \mathbb{Y}_n is η_{kn}

Thus, according to Inference Rule 1 and 2, we can infer that $\forall j(j = 1, 2, \dots, n)$ are

$$\eta_j^* = \sum_{i=1}^k \frac{c_i \cdot \eta_{ij} |_{(a_1 \in \xi_{i1}) \cap (a_2 \in \xi_{i2}) \cap \dots \cap (a_m \in \xi_{im})}}{c_1 + c_2 + \dots + c_k},$$

where $c_i = \mathcal{M}\{(a_1 \in \xi_{i1}) \cap (a_2 \in \xi_{i2}) \cap \dots \cap (a_m \in \xi_{im})\}$ for $i = 1, 2, \dots, k$. Then, by using the expected value operator, we obtain

$$\beta_j = E[\eta_j^*]$$

for $j = 1, 2, \dots, n$. Now, we construct a function from crisp inputs $\alpha_1, \alpha_2, \dots, \alpha_m$ to crisp outputs $\beta_1, \beta_2, \dots, \beta_n$, i.e.,

$$(\beta_1, \beta_2, \dots, \beta_n) = f(\alpha_1, \alpha_2, \dots, \alpha_m).$$

Then, we get an uncertain system f . For the uncertain system we proposed, we have the following theorem.

Theorem 4. [13] *Assume that $\xi_{i1}, \xi_{i2}, \dots, \xi_{im}$ and $\eta_{i1}, \eta_{i2}, \dots, \eta_{in}$ are independent uncertain sets with membership functions $\mu_{i1}, \mu_{i2}, \dots, \mu_{im}, \nu_{i1}, \nu_{i2}, \dots, \nu_{in}$, $i = 1, 2, \dots, k$, respectively. Then, the uncertain system from $\alpha_1, \alpha_2, \dots, \alpha_m$ to $\beta_1, \beta_2, \dots, \beta_n$ is*

$$b_j = \sum_{i=1}^k \frac{c_i \cdot E[\eta_{ij}^*]}{c_1 + c_2 + \dots + c_k},$$

where $j = 1, 2, \dots, n$ and η_{ij}^* are uncertain sets whose membership functions are given by

$$\nu_{ij}^*(y) = \begin{cases} \frac{\nu_{ij}(y)}{c_i}, & \text{if } \nu_{ij}(y) < \frac{c_i}{2} \\ \frac{\nu_{ij}(y) + c_i - 1}{\mu(a)}, & \text{if } \nu_{ij}(y) > 1 - \frac{c_i}{2} \\ 0.5, & \text{otherwise} \end{cases}$$

and $c_i = \min_{1 \leq l \leq m} \mu_{il}(a_l)$ are constants.

Next, we discuss the expected value of a special triangular uncertain set. Without loss of generality, we assume $n = 1$. Then the uncertain system proposed in the above becomes:

$$b = \sum_{i=1}^k \frac{c_i \cdot E[\eta_i^*]}{c_1 + c_2 + \dots + c_k}, \tag{5}$$

$$\nu_i^*(y) = \begin{cases} \frac{\nu_i(y)}{c_i}, & \text{if } \nu_i(y) < \frac{c_i}{2} \\ \frac{\nu_i(y) + c_i - 1}{\mu(a)}, & \text{if } \nu_i(y) > 1 - \frac{c_i}{2} \\ 0.5, & \text{otherwise,} \end{cases} \tag{6}$$

$$c_i = \min_{1 \leq l \leq m} \mu_{il}(a_l). \tag{7}$$

Theorem 5. *Assume we have an uncertain system with m inputs and 1 output consisting of k inference rules. The antecedents of the rules are represented by the uncertain sets ξ_i with membership functions $\mu_{i1}, \mu_{i2}, \dots, \mu_{im}$, $i = 1, 2, \dots, k$. And the consequent is represented by an triangular uncertain set $\eta_i = (\alpha_i, \beta_i, \gamma_i)$ with a membership function ν_i , where*

the coefficients satisfy

$$\alpha_i + \gamma_i = 2\beta_i, \quad i = 1, 2, \dots, k. \tag{8}$$

We have

$$E[\eta_i^*] = \beta_i, \quad i = 1, 2, \dots, k.$$

Proof. Given the m input data a_1, a_2, \dots, a_m , we can calculate c_i from Equation 7. Then, we can get the membership functions v_i^* of the consequence uncertain sets η_i^* according to Equation 6. Next, the computation of the expected value of uncertain consequence breaks into three cases.

Case 1: Assume $c_i/2 = 0.5$. We can immediately have $v_i^*(y) = v_i(y)$, thus

$$E[\eta_i^*] = \frac{\alpha_i + 2\beta_i + \gamma_i}{4} = \beta_i.$$

Case 2: Assume $c_i/2 < 0.5$. Let y_{11}^i and y_{12}^i ($y_{11}^i < y_{12}^i$) be the two points that satisfy the equation $v_i(y) = c_i/2$. Similarly, y_{21}^i and y_{22}^i ($y_{21}^i < y_{22}^i$) satisfy the equation $v_i(y) = 1 - c_i/2$. Since the membership function of a triangular uncertain set has a symmetry property, we have

$$y_{11}^i + y_{12}^i = 2\beta_i, \quad y_{21}^i + y_{22}^i = 2\beta_i. \tag{9}$$

Then, we can rewrite the membership function of η_i as follows:

$$v_i^* = \begin{cases} \frac{v_i(y)}{c_i}, & \text{if } \alpha_i \leq y < y_{11}^i \\ \frac{v_i(y)+c_i-1}{c_i}, & \text{if } y_{21}^i \leq y < y_{22}^i \\ \frac{v_i(y)}{c_i}, & \text{if } y_{12}^i \leq y < \gamma_i \\ 0.5, & \text{otherwise.} \end{cases} \tag{10}$$

And $v_i^*(\beta_i) = 1$. Together with Equations 3, 8, and 9, we have

$$\begin{aligned} E[\eta_i^*] &= \beta_i + \frac{1}{2} \left(\int_{\beta_i}^{y_{22}^i} \frac{v_i(y) + c_i - 1}{c_i} dy + \int_{y_{22}^i}^{y_{12}^i} 0.5 dy + \int_{y_{12}^i}^{\gamma_i} \frac{v_i(y)}{c_i} dy \right) \\ &\quad - \frac{1}{2} \left(\int_{y_{21}^i}^{\beta_i} \frac{v_i(y) + c_i - 1}{c_i} dy + \int_{y_{11}^i}^{y_{21}^i} 0.5 dy + \int_{\alpha_i}^{y_{11}^i} \frac{v_i(y)}{c_i} dy \right) \\ &= \beta_i + \frac{1}{2} \left(\int_{\beta_i}^{y_{22}^i} \frac{v_i(y) + c_i - 1}{c_i} dy - \int_{y_{21}^i}^{\beta_i} \frac{v_i(y) + c_i - 1}{c_i} dy \right) \\ &\quad + \frac{1}{2} \left(\int_{y_{12}^i}^{\gamma_i} \frac{v_i(y)}{c_i} dy - \int_{\alpha_i}^{y_{11}^i} \frac{v_i(y)}{c_i} dy \right) \\ &\quad + \frac{1}{2} \left(\int_{y_{22}^i}^{y_{12}^i} 0.5 dy - \int_{y_{11}^i}^{y_{21}^i} 0.5 dy \right) \\ &= \beta_i. \end{aligned}$$

Case 3: Assume $c_i > 0.5$. Similarly, we have $E[\eta_i^*] = \beta_i$. Thus, we have proved the theorem. \square

Problem formulation

In this section, we propose an extraction model to obtain uncertain inference rules.

Let $X = (x_1, x_2, \dots, x_n)$ be the decision vector, which represents a rule base consisting of n rules. Each rule has m antecedents which are described by Q uncertain sets and one consequent which is described by R uncertain sets. Each variable x_i represents a sequence $x_{i1}x_{i2} \dots x_{im}x_{im+1}$, where $x_{ij} \in \{0, 1, 2, \dots, Q\} (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ represent the antecedents of the inference rule. And $x_{im+1} \in \{0, 1, 2, \dots, R\} (i = 1, 2, \dots, n)$ represent the consequent. Thus, each variable of decision vector represents one inference rule. Some $x_{ij} = 0$ means this antecedent is not included. And some $x_{im+1} = 0$ means this inference rule will not be included in the rule base. For example, assume that we have one inference rule consists of 4 antecedents and 1 consequent. They are described by 5 uncertain sets which refer to five descriptions: very low, low, medium, high, and very high. We use 1, 2, 3, 4, 5 to denote them. Thus, sequence "23045", for example, represents the rule: "if input 1 is low, input 2 is medium, and input 4 is high, then the output is very high".

Uncertain systems can be used for classification. But which uncertain system is better depends on the rule base. Here, we try to find best rule base by comparing the mean absolute errors of the origin output and the system output. That is,

$$MAE = \frac{1}{P} \sum_{i=1}^P |o_i - t_i|, \tag{11}$$

where P is the number of training data, $o_i, t_i (i = 1, 2, \dots, P)$ are the system outputs and origin outputs, respectively. If we find the rule base with the least mean absolute error, we extract the uncertain inference rules successfully. We can obtain the system outputs by Equation 5. However, they may not be integers. To avoid this nonsense, for a classification problem with C classes, we can divide interval that covers all the system outputs into C subintervals. Then, if the output from Equation 5 is in the i th subinterval, we have $o_i = i$. Thus, we transfer the classification problem to the following optimization model:

$$\left\{ \begin{array}{l} \min_X F(X) = MAE \\ s.t. \\ X = (x_1, x_2, \dots, x_n) \\ x_i = x_{i1} \dots x_{im}x_{im+1} \\ x_{ij} \in \{0, 1, \dots, Q\} \\ x_{im+1} \in \{0, 1, \dots, R\} \\ i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{array} \right.$$

Extraction method for uncertain inference rules with mutations

In this section, we propose the extraction method for uncertain inference rules with mutations by ant colony optimization algorithm.

As stated before, each x_i is a sequence of m values in $\{0, 1, 2, \dots, Q\}$ and 1 value in $\{0, 1, 2, \dots, R\}$. Without loss of generality, we set $Q = R$. Each number in $\{0, 1, 2, \dots, Q\}$ is a node. Let ants walking across these nodes. Ants choose the next node in probability based on the pheromone levels in the $Q + 1$ choices at every step. Once ants move $m + 1$ steps, a candidate decision variable is generated. After repeat this process n times, we get a candidate solution. After all ants finish their walk, update the pheromone trails. Denote the pheromone trail by $\tau_{i,k,j}(t)$ associated to the node j at step k of x_i in iteration t . The procedures are described as follows.

(1) **Initialization:** Randomly generate a feasible solution X_0 , and set the optimal solution $\hat{X} = X_0$. Set $\tau_{i;k,j}(0) = \tau_0, i = 1, 2, \dots, n, k = 1, 2, \dots, m + 1, j = 0, 1, 2, \dots, Q$, where τ_0 is a fixed parameter.

(2) **Ant movement:** At each step k after building the sequence $x_{i1}x_{i2} \dots x_{ik}$, select the next node in probability following

$$p_{k;k+1} = \frac{\tau_{i;k+1,j}(t)}{\sum_{q=0}^Q \tau_{i;k+1,q}(t)}. \tag{12}$$

In this way, we could get a sequence $x_{i1}x_{i2} \dots x_{im+1}$. To speed up the algorithm, we mutate this sequence to get a new candidate sequence. The mutation is made as follows: randomly add 1 or subtract 1 to each element x_{ij} in the sequence; if the element is 0, the mutated element is 1; if the element is Q , the mutated element is $Q - 1$. Assume X' is the mutated solution, if $\Delta F = F(X') - F(X) \leq 0$, then $X \leftarrow X'$; otherwise, keep the current solution. If Q is very large, we could repeat this mutation until some termination condition is satisfied.

(3) **Pheromone Update:** At each iteration t , let \hat{X} be the optimal solution found so far and X_t be the best feasible solution in the current iteration. Assume $F(\hat{X})$ and $F(X_t)$ are the corresponding objective function values.

If $F(X_t) < F(\hat{X})$, then $\hat{X} \leftarrow X_t$.

Reinforce the pheromone trails on nodes of \hat{X} and evaporate the pheromone trails on the left nodes:

$$\tau_{ij,k}(t) = \begin{cases} (1 - \rho)\tau_{ij,k}(t - 1) + \rho g(\hat{X}), & \text{if } (k, j) \in \hat{X} \\ (1 - \rho)\tau_{ij,k}(t - 1), & \text{otherwise} \end{cases} \tag{13}$$

where ρ ($0 < \rho < 1$) is the evaporation rate, $g(x)$ ($0 < g(x) < +\infty$) is a function with that $g(x) \geq g(y)$ if $F(x) < F(y)$, for example, $g(x) = L/(|F(x)| + 1)$ is a function satisfying the condition where $L > 0$.

Let τ_0 be the initial value of pheromone trails, n be the number of decision variables, M be the number of ants, ρ be evaporation rate and T be the number of iterations. Now, we summarize the algorithm as follows.

Step 1 Initialize all pheromone trails with the same pheromone level τ_0 . Randomly generate a feasible solution X_0 , and set optimal solution $\hat{X} = X_0$. Set $l \leftarrow 0$.

Step 2 Ant movement in probability following Equation 12. Generate a decision variable x_i after $m + 1$ steps.

Step 3 Repeat Step 2 until $X = (x_1, x_2, \dots, x_n)$ is generated; mutate every x_i : thus, generate a new decision vector $X' = (x'_1, x'_2, \dots, x'_n)$; if $\Delta F = F(X') - F(X) \leq 0$, then $X \leftarrow X'$.

Step 4 Repeat Step 2 and Step 3 for all M ants.

Step 5 Calculate the system outputs by Equation 5. Then, calculate the objective function values for the M candidate solutions by Equation 11. Denote the best solution in this iteration by X_l .

Step 6 If $F(X_l) < F(\hat{X})$, then $\hat{X} \leftarrow X_l$; update the pheromone trails according to Equation 13.

Step 7 $l \leftarrow l + 1$; if $l = T$, terminate; otherwise, go to Step 2.

Step 8 Report the optimal solution \hat{X} .

With this algorithm above, we obtain an uncertain rule base. Then, we successfully design an uncertain system and can use it for classification.

Extraction method for uncertain inference rules with SA

In the previous section, to speed up the algorithm, we introduce a mutation operation. Here, we introduce the simulated annealing algorithm as the local search operation.

Simulated annealing algorithm was initiated by Metropolis in 1953, applied to portfolio optimization by Kirkpatrick [25] in 1983. The name and inspiration come from annealing in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. Simulated annealing algorithm is excellent at avoiding getting stuck in local optimums. It has a good robust property and is universal and easy to implement.

For optimization problem (1), we can use simulated annealing algorithm to search for the optimal solution. The algorithm is as follows.

Step 1 Randomly generate a initial solution x_0 ; $x \leftarrow x_0$; $k \leftarrow 0$; $t_0 \leftarrow t_{max}$ (initial temperature);

Step 2 If the temperature satisfies the inner cycle termination criterion, go to Step 3; otherwise, randomly choose a point x' in the neighborhood $N(x)$, calculate $\Delta f = f(x') - f(x)$. If $\Delta f \leq 0$, then $x \leftarrow x'$; otherwise, according to Metropolis acceptance criterion, if $\exp(-\Delta f/t_k) > \text{random}(0, 1)$, then $x \leftarrow x'$. Repeat Step 2.

Step 3 $t_{k+1} = d(t_k)$ (temperature decrease); $k \leftarrow k + 1$; if the termination criterion is satisfied, stop and report the optimal solution; otherwise, go to Step 2.

In this section, we combine ant colony optimization algorithm and simulated annealing algorithm. In each iteration of ant colony optimization algorithm, we get a feasible solution. Then, we use it as the initial solution of the simulated annealing algorithm to get a neighbor solution. This neighbor solution will be accepted in probability. And for each decision vector $X = (x_1, x_2, \dots, x_n)$, $x_i = x_{i1}x_{i2} \dots x_{im+1}$, we build the neighbor solution as follows: for each x_i , for some randomly generated p and q ($1 \leq p < q \leq m$), reverse the order of the sequence $x_{ip} \dots x_{iq}$, i.e., $x'_i = x_{i1} \dots x_{ip-1}x_{iq}x_{iq-1} \dots x_{ip+1}x_{ip}x_{iq+1} \dots x_{im+1}$. For example, assume x_i is 0123456, $p = 2$, $q = 6$, and the neighbor solution x'_i is 0543216. In this way, we obtain a neighbor solution X' . If $\Delta F = F(X') - F(X) \leq 0$, $X \leftarrow X'$; otherwise, if $\exp(-\Delta F/t_k) > \text{random}(0, 1)$, then $X \leftarrow X'$; otherwise, abandon this neighbor solution. Still denote the pheromone trail by $\tau_{i;k,j}(t)$. The procedure are described as follows.

(1) **Initialization:** Generate a feasible solution X_0 randomly and set the optimal solution $\hat{X} = X_0$. Set $\tau_{i;k,j}(0) = \tau_0$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, m + 1$, $j = 0, 1, 2, \dots, Q$, where τ_0 is a fixed parameter.

(2) **Ant movement:** At each step k after building the sequence $x_{i1}x_{i2} \dots x_{ik}$, select the next node in probability following Equation 12. In this way, we could get a sequence $x_{i1}x_{i2} \dots x_{im+1}$. In order to expand the search range, we use simulated annealing algorithm to search locally around the solution at this step. Assume the neighbor solution is X' . If $\Delta F = F(X') - F(X) \leq 0$, $X \leftarrow X'$; otherwise, if $\exp(-\Delta F/t_k) > \text{random}(0, 1)$ where t_k is the current temperature and $t_k \rightarrow 0$ when $k \rightarrow \infty$, then $X \leftarrow X'$; otherwise, abandon this neighbor solution and still choose the original feasible solution.

(3) **Pheromone Update:** Let \hat{X} be the optimal solution found so far and X_t be the best feasible solution in the current iteration t . Assume $F(\hat{X})$ and $F(X_t)$ are the corresponding objective function values. To avoid the optimal solution \hat{X} getting stuck in local optimums, we also use acceptance function here.

If $F(X_t) < F(\hat{X})$, then $\hat{X} \leftarrow X_t$.

Build a neighbor solution \hat{X}' .

If $F(\hat{X}') \leq F(\hat{X})$, then $\hat{X} \leftarrow \hat{X}'$;

If $F(\hat{X}') > F(\hat{X})$, check the Metropolis acceptance criterion, i.e., if $\exp(-\Delta\hat{F}/T_t) > \text{random}(0, 1)$, $T_t \rightarrow 0$, $t \rightarrow \infty$, then $X^* \leftarrow \hat{X}'$.

Reinforce the pheromone trails on the nodes of \hat{X} and X^* and evaporate the pheromone trails on the left nodes:

$$\tau_{ij,k}(t) = \begin{cases} (1 - \rho)\tau_{ij,k}(t - 1) + \rho g(\hat{X}), & \text{if } (k, j) \in \hat{X} \\ (1 - \rho)\tau_{ij,k}(t - 1) + \frac{\rho}{2}g(\hat{X}), & \text{if } (k, j) \in X^* \\ (1 - \rho)\tau_{ij,k}(t - 1), & \text{otherwise} \end{cases} \tag{14}$$

where, ρ ($0 < \rho < 1$) is the evaporate rate, and $g(x)$ ($0 < g(x) < +\infty$) is a function with that $g(x) \geq g(y)$ if $F(x) < F(y)$. For example, $g(x) = L/(|F(x)| + 1)$ is an available function if $L > 0$.

Now, we summarize the algorithm as follows.

Step 1 Initialize all pheromone trails with the same pheromone level τ_0 . Randomly generate a feasible solution X_0 , and set optimal solution $\hat{X} = X_0$. Set $t \leftarrow 0$.

Step 2 Ant movement in probability following Equation 12. Generate a decision variable x_i after $m + 1$ steps.

Step 3 Repeat Step 2 until decision vector $X = (x_1, x_2, \dots, x_n)$ is generated. Build the neighbor solution X' . If $\Delta F = F(X') - F(X) \leq 0$, $X \leftarrow X'$; otherwise, if $\exp(-\Delta F/t_k) > \text{random}(0, 1)$ where t_k is the current temperature and $t_k \rightarrow 0$ when $k \rightarrow \infty$, then $X \leftarrow X'$.

Step 4 Repeat Step 2 and Step 3 until all ants finish their walk, and generate M candidate solutions.

Step 5 Calculate the system outputs by Equation 5. Then, calculate the objective function values for the M candidate solutions by Equation 11. Denote the best solution in this iteration by X_t .

Step 6 If $F(X_t) < F(\hat{X})$, then $\hat{X} \leftarrow X_t$. Build the neighbor solution of \hat{X} , which is denoted by \hat{X}' . If $\Delta\hat{F} = F(\hat{X}') - F(\hat{X}) \leq 0$, then $\hat{X} \leftarrow \hat{X}'$; otherwise, if Metropolis acceptance criterion is satisfied, i.e., if $\exp(-\Delta\hat{F}/T_t) > \text{random}(0, 1)$, $T_t \rightarrow 0$, $t \rightarrow \infty$, then $X^* \leftarrow \hat{X}'$.

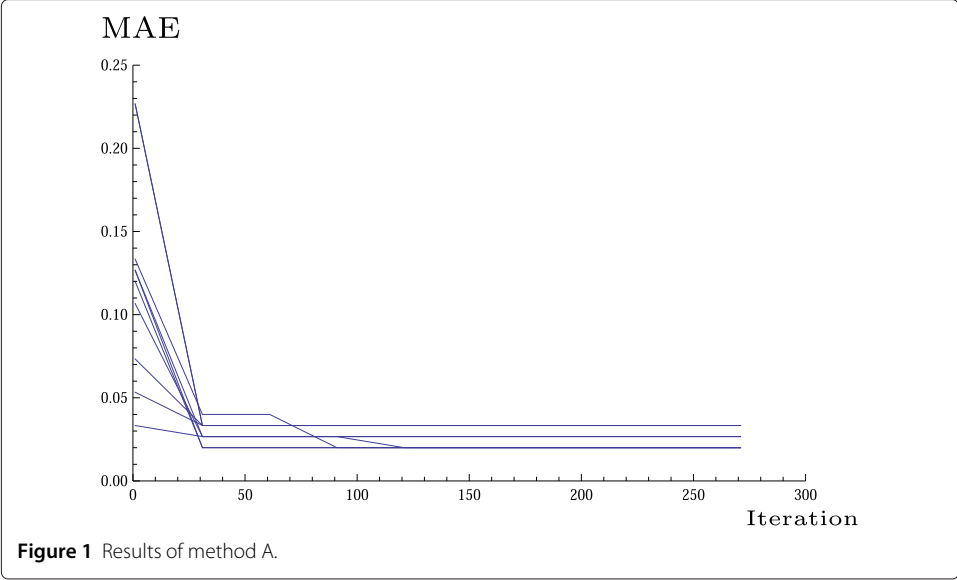
Step 7 Update the pheromone trails according to Equation 14.

Step 8 $t \leftarrow t + 1$; if $t = T$, terminate; otherwise, go to Step 2.

Step 9 Report the optimal solution \hat{X} .

Table 1 Parameters

	a_p	b_p	c_p
$\rho = 1$	0.5	1.01	1.52
$\rho = 2$	1.7	2.74	4.48
$\rho = 3$	5	6.07	7.14



Experiments

In this section, we use our two extraction methods to extract uncertain inference rules. And then use the uncertain systems to solve some classification problems. We applied our methods to the IRIS [26] classification problem and the Wisconsin Breast Cancer (WBC) [27] classification problem.

IRIS classification

IRIS data set is the typical date set in data classification. It contains 150 instances of 3 classes, which are Setosa, Versicolor, and Virginica. Each class has 50 instances. Each instance has 4 attributes which are sepal length (SL), sepal width (SW), petal length (PL),

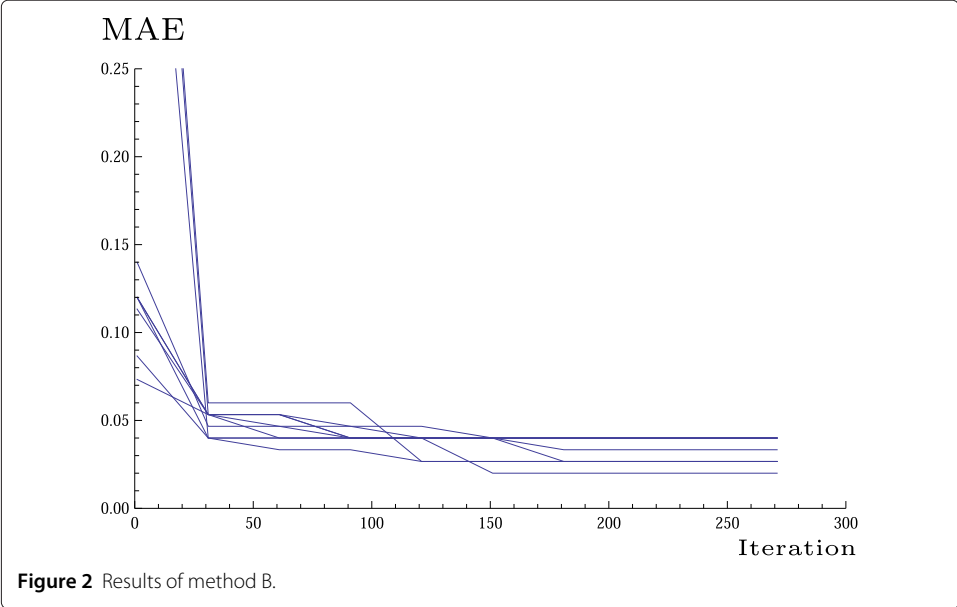


Table 2 Accuracy comparison

Method	Paper	Accuracy rate (%)
Method A	This paper	97.33
Method B	This paper	97.5
C4.5	[2]	94.0
ACOA	[5]	96.6
MACO	[6]	95.53
HNFAQ	[28]	98.67

Table 3 IRIS classification rules extracted by method A

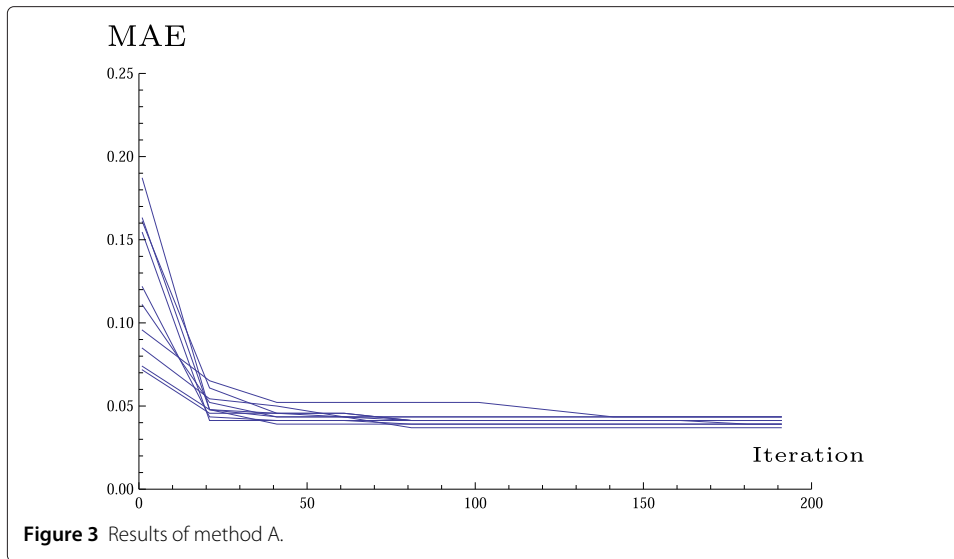
IF				THEN
SL	SW	PL	PW	Class
1	3	1	3	1
1	0	1	1	1
1	2	3	2	1
1	1	2	1	2
2	1	0	3	2
3	2	0	2	3
1	1	3	3	3

Table 4 IRIS classification rules extracted by method B

IF				THEN
SL	SW	PL	PW	Class
3	2	3	1	2
1	1	0	0	2
0	2	1	1	3
0	1	1	3	1
1	1	3	3	2
1	1	3	1	1
2	1	1	2	1

Table 5 Parameters

	a_p	b_p	c_p
$p = 1$	0.3	1.01	1.72
$p = 2$	2	6.07	10.14



and petal width (PW). They are described by 3 uncertain sets: low (1), medium (2), and high (3). The membership functions are

$$\mu_q(x) = \exp\left(-\frac{(x - V_q)^2}{2\beta^2}\right),$$

where x is the input, $\beta = 0.618$ and $V_q = \frac{q-1}{2}$, $q = 1, 2, 3$. Based on these 4 attributes, we try to infer which class does the instance belong to. We use 3 triangular uncertain sets $\eta_p = (a_p, b_p, c_p)$ ($p = 1, 2, 3$) to describe the possible classes (class 1: Setosa; class 2: Versicolor; class 3: Virginica). And the parameters $a_p, b_p, c_p \in \mathcal{R}$ are listed in Table 1.

First, we normalize the data to $[0, 1]$ to simplify the computation. IRIS data set is our training set while it is also used for testing. Then, we set maximum number of rules $n = 10$, number of ants $M = 10$, evaporate rate $\rho = 0.3$, and number of iterations $T = 300$. Each algorithm runs ten times. The results are in Figures 1 and 2. Denote the extraction

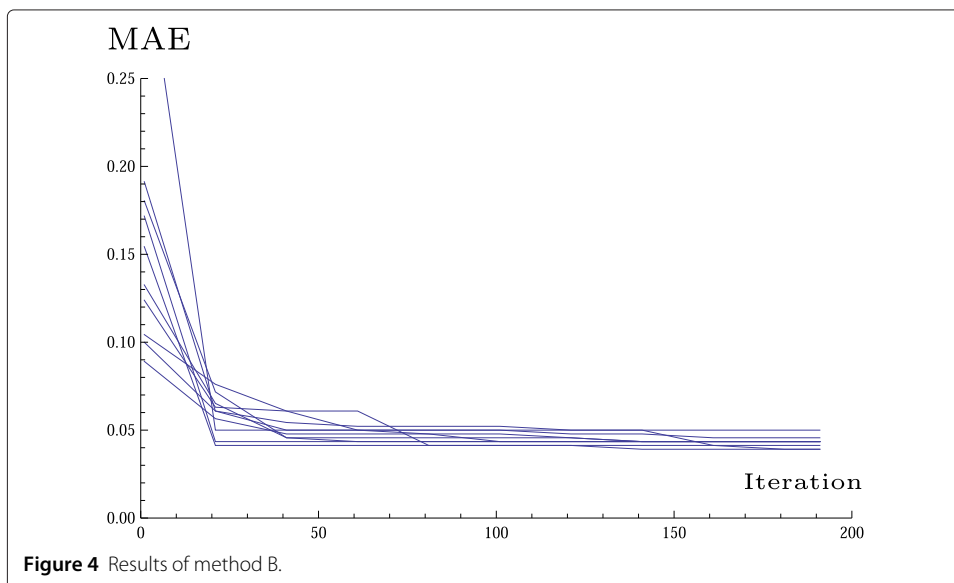


Table 6 Accuracy rate comparison

Method	Paper	Accuracy rate (%)
Method A	This paper	98.3
Method B	This paper	98.33
C4.5	[2]	94.25
ACOA	[5]	97.91
MACO	[6]	97.07
FMM	[29]	97.86

method with mutation by A and the method with SA by B. It can be seen that the method A converges fast at about 120th iteration. And method B converges a little slower at about 150th iteration.

Then, we can classify the IRIS data with the uncertain systems we introduced. We find the average accuracy rates of the two methods are 97.33% and 97.5%, respectively. Comparison with other methods are listed in Table 2.

List the rule bases we get with the highest accuracy rates (98.0% and 98.67%, respectively) in Tables 3 and 4. Note that although the maximum number of rules is 10, the final rule bases we obtain has only 7 rules.

Wisconsin Breast Cancer classification

Wisconsin Breast Cancer data set is a common medical data set. It contains 699 instances of 2 classes, which are sick and healthy. Two hundred forty-one instances are sick and 458 instances are healthy. Each instances has 9 attributes, which are clump thickness (CT), uniformity of cell size (UCS), uniformity of cell shape (UCCS), marginal adhesion (MA), single epithelial cell size (SPCS), bare nuclei (BN), bland chromatin (BC), normal nucleoli (NN), and mitoses (MT). They are described by 5 uncertain sets: very low (1), low (2), medium (3), high (4), and very high (5). The membership functions are

$$\mu_q(x) = \exp\left(-\frac{(x - V_q)^2}{2\beta^2}\right),$$

where x is the input, $\beta = 0.4247$, and $V_q = \frac{q-1}{2}$, $q = 1, 2, 3, 4, 5$. Based on these attributes, we diagnose whether one instance is sick or not. We use 2 triangular uncertain sets $\eta_p = (a_p, b_p, c_p)$ ($p = 1, 2$) to describe the possible classes (sick and healthy). And the parameters $a_p, b_p, c_p \in \mathcal{R}$ are listed in Table 5.

Table 7 WBC classification rules extracted by method A

IF									THEN
CT	UCS	UCCS	MA	SPCS	BN	BC	NN	MT	Class
1	5	0	3	2	0	4	2	1	2
1	2	1	1	4	4	2	1	0	1
1	3	5	2	3	2	1	1	4	2
1	3	4	2	3	1	2	2	1	1
2	4	4	1	1	2	4	5	1	2
3	3	4	5	4	3	2	4	4	2
5	2	4	0	3	0	0	2	1	1
2	4	4	1	1	2	4	5	1	1
2	4	2	3	5	3	2	5	5	2

Table 8 WBC classification rules extracted by method B

IF									THEN
CT	UCS	UCCS	MA	SPCS	BN	BC	NN	MT	Class
5	3	3	3	3	2	3	2	4	2
0	0	0	0	0	0	0	4	0	1
4	4	4	4	0	1	1	1	4	2
1	4	4	1	1	2	1	5	1	2
1	1	1	2	0	3	5	5	5	2
0	3	0	0	0	0	0	1	2	1

First, we normalize the data to $[0, 1]$ to simplify the computation. The first 460 instances are used for training while the left 239 instances are used for testing. Then, we set maximum number of rules $n = 10$, number of ants $M = 20$, evaporate rate $\rho = 0.3$, and number of iterations $T = 200$. Each algorithm runs ten times. The results are in Figures 3 and 4. We still find that method A converges faster than method B. Method A stabilizes at about 50th iteration while method B stabilizes until about 80th iteration.

Then, we test the uncertain systems we get with the later 239 instances. We find the average accuracy rates of the two methods on the training set are 96.0% and 96.26%, respectively. Using the uncertain system with the highest accuracy rate of each method on the test set, we find the accuracy rates are 98.37% and 98.33%. Comparison with other methods are listed in Table 6.

The rule base with the highest accuracy rates (98.37% and 98.33%, respectively) on the test set are listed in Tables 7 and 8. Method A gives us a rule base of 9 rules, and method B provides a rule base of 6 rules.

We apply our two extraction methods to the classification problems of IRIS data set and WBC data set. Compare our results with other researchers' work, we can find that both methods have higher accuracy rate than ACOA and MACO in two classification problems. And for IRIS data set, accuracy rates of method A and B are lower than HNFQ but higher than C4.5. For WBC data set, their accuracy rates are higher than C4.5 and FMM.

Conclusions

In this paper, we designed an uncertain system for data classification. And we proposed two extraction methods for uncertain inference rules by using ant colony optimization algorithm. Then, we applied our methods to IRIS classification problem and WBC classification problem. Our methods are shown to be superior in accuracy to some existing methods.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.61273009).

Received: 14 December 2014 Accepted: 20 April 2015

Published online: 14 May 2015

References

1. Kantardzic, M: Data Mining: Concepts, Models, Methods, and Algorithms. 2nd ed. Wiley, Hoboken (2011)
2. Quinlan, JR: Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.* **4**(1), 77–90 (1996)
3. Parpinelli, RS, Lopes, HS, Freitas, AA: Data mining with an ant colony optimization algorithm. *IEEE Trans. Evolut. Comput.* **6**(4), 321–332 (2002)
4. Casillas, J, Cordon, O, Herrera, F: Learning fuzzy rules using ant colony optimization algorithms. In: Proceedings of the 2nd International Workshop on Ant Algorithms: From Ant Colonies to Artificial Ants, pp. 13–21, Brussels, (2000)

5. Zhu, Y: Ant colony optimization-based hybrid intelligent algorithms. *World J. Modell. Simul.* **2**(5), 283–289 (2006)
6. Zhu, Y: An intelligent algorithm: MACO for continuous optimization models. *J. Intell. Fuzzy Syst.* **24**, 31–36 (2013)
7. Lee, Z, Su, S, Chuang, C, Liu, K: Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment. *Appl. Soft Comput.* **8**(1), 55–78 (2008)
8. Shelokar, PS, Siarry, P, Jayaraman, VK, Kulkarni, BD: Particle swarm and ant colony algorithms hybridized for improved continuous optimization. *Appl. Math. Comput.* **188**(1), 129–142 (2007)
9. Liu, B: *Uncertainty Theory*. 2nd ed. Springer, Berlin (2007)
10. Liu, B: Fuzzy process, hybrid process and uncertain process. *J. Uncertain Syst.* **2**(1), 3–16 (2008)
11. Chen, X, Liu, B: Existence and uniqueness theorem for uncertain differential equations. *Fuzzy Optimization Decis. Mak.* **9**(1), 69–81 (2010)
12. Liu, B: Some research problems in uncertainty theory. *J. Uncertain Syst.* **3**(1), 3–10 (2009)
13. Liu, B: *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*. Springer, Berlin (2010)
14. Peng, J, Yao, K: A new option pricing model for stocks in uncertainty markets. *Int. J. Oper. Res.* **8**(2), 18–26 (2011)
15. Zhu, Y: Uncertain optimal control with application to a portfolio selection model. *Cybern. Syst.* **41**(7), 535–547 (2010)
16. Liu, B: Uncertain set theory and uncertain inference rule with application to uncertain control. *J. Uncertain Syst.* **4**(2), 83–98 (2010)
17. Liu, B: Uncertain logic for modeling human language. *J. Uncertain Syst.* **5**(1), 3–20 (2011)
18. Gao, X, Gao, Y, Ralescu, DA: On Liu's inference rule for uncertain systems. *Int. J. Uncertain. Fuzz. Knowledge-Based Syst.* **18**(1), 1–11 (2010)
19. Peng, Z, Chen, X: Uncertain systems are universal approximators. *J. Uncertainty Anal. Appl.* **2**, Article, 13 (2014)
20. Gao, Y: Uncertain inference control for balancing inverted pendulum. *Fuzzy Optimization Decis. Mak.* **11**(4), 481–492 (2012)
21. Liu, B: Membership functions and operational law of uncertain sets. *Fuzzy Optimization Decis. Mak.* **11**(4), 387–410 (2012)
22. Zadeh, LA: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man Cybern.* **3**(1), 28–44 (1973)
23. Mamdani, EH: Applications of fuzzy algorithms for control of a simple dynamic plant. *Proc. Institution Electr. Eng. Control Sci.* **121**(12), 1585–1588 (1974)
24. Takagi, K, Sugeno, M: Fuzzy identification of system and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **15**(1), 116–132 (1985)
25. Kirkpatrick, S, Gelatt, CD, Vecchi, MP: Optimization by simulated annealing. *Science.* **220**(4598), 671–680 (1983)
26. Iris dataset (1936). <https://archive.ics.uci.edu/ml/datasets/Iris>
27. Wisconsin Breast Cancer Dataset (1992). [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Original\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original))
28. de Souza, FJ, Vellasco, M, Pacheco MA: Hierarchical neuro-fuzzy quadtree models. *Fuzzy Sets Syst.* **130**(2), 189–205 (2002)
29. Gabrys, B, Bargiela, A: General fuzzy min-max neural network for clustering and classification. *IEEE Trans. Neural Networ.* **11**(3), 769–783 (2000)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
