# An Innovative Method to Classify Remote-Sensing Images Using Ant Colony Optimization

Xiaoping Liu, Xia Li, Lin Liu, Jinqiang He, and Bin Ai

***Abstract*—This paper presents a new method to improve the classification performance for remote-sensing applications based on swarm intelligence. Traditional statistical classifiers have limitations in solving complex classification problems because of their strict assumptions. For example, data correlation between bands of remote-sensing imagery has caused problems in generating satisfactory classification using statistical methods. In this paper, ant colony optimization (ACO), based upon swarm intelligence, is used to improve the classification performance. Due to the positive feedback mechanism, ACO takes into account the correlation between attribute variables, thus avoiding issues related to band correlation. A discretization technique is incorporated in this ACO method so that classification rules can be induced from large data sets of remote-sensing images. Experiments of this ACO algorithm in the Guangzhou area reveal that it yields simpler rule sets and better accuracy than the See 5.0 decision tree method.**

***Index Terms*—Ant colony optimization (ACO), artificial intelligence (AI), classification, remote sensing.**

## I. Introduction

CLASSIFICATION, which extracts useful information from remote-sensing data, has been one of the key topics in remote-sensing studies [1], [2]. For example, classification is frequently carried out to obtain land use/cover information. Local, regional, and global environmental changes are closely related to land use/cover and its changes over time [3]. Remote-sensing imagery has been an important source of acquiring land use/cover information. Numerous methods for remote-sensing classification have been developed in the last three decades, including statistical classifiers, knowledge-based systems (KBS), neural networks, and other artificial intelligence (AI) methods [4]. However, these methods still have limitations because of the complexities of remote-sensing classification. For example, maximum-likelihood classifiers, the most commonly used statistical method, perform classification according to the likelihood between unknown and known pixels in the sample data [5], [6]. This method requires that each class data approximately follow a normal distribution. The quality of training samples, which are used to estimate the parameters of the classifier, is key to the overall accuracy of the classification. Another technique, i.e., KBS, usually integrates spectral information, spatial structures, and experts' experiences for classification [7]–[9]. KBS are applicable in most situations because they do not require data to follow a normal distribution. Their classification rules are also easy to understand, and the classification processes are analogous to human reasoning. KBS can improve the accuracy of classification relative to statistical classifiers in some situations. However, being constrained by the long and repeated process of acquiring and assimilating knowledge from experts' experience, this type of classification has not been widely applied [2]. Neural network classifiers, being self-adaptive, are ideal for complicated and parallel computation [10]–[12], but the derived classification rules are difficult, if not impossible, to interpret. In addition, they are prone to overfitting the training set, getting trapped in the local optima, and slowly converging to a solution.

Recent technologies and theories of AI provide new opportunities for efficient classifications of remote-sensing data [13]. Ant intelligence, for example, has been used to solve complex classification problems. Ant colony optimization (ACO), a computational method derived from natural biological systems, was first proposed by Colorni *et al.* in 1991 [14]. ACO is a computer optimization algorithm that simulates the behaviors of real ants in their search for the shortest paths to food sources. ACO looks for optimal solutions by utilizing distributed computing, local heuristics, and knowledge from past experience [15]. The main characteristic of ACO is the utilization of indirect communication by ants through laying pheromone along their routes. Another characteristic is the positive feedback mechanism that facilitates the rapid discovery of optimal solutions [16], [17]. Thus, ACO is essentially a complex multiagent system where low-level interactions between individual agents result in complex behavior of the whole ant colony. This method has proven to be robust and versatile [18].

Satisfactory results have been obtained in solving traveling salesman problems, data clustering, combinatorial optimization, and network routing by using ACO [19]–[22]. However, there are very limited studies on classification rule induction using ACO, although ACO is very successful in global search and can better cope with attribute correlation than other rule induction algorithms such as a decision tree [23]. Parpinelli *et al.* [24], [25] were the first to propose ACO for discovering classification rules using a system called Ant-Miner. Their study demonstrates that Ant-Miner produces better accuracy

Manuscript received November 12, 2006; revised April 3, 2007, August 2, 2007, December 25, 2007, and March 31, 2008. Current version published November 26, 2008. This work was supported by the National Outstanding Youth Foundation of China under Grant 40525002, the Key National Natural Science Foundation of China under Grant 40830532, and the Hi-Tech Research and Development Program of China (863 Program) under Grant 2006AA12Z206.

X. Liu, J. He, and B. Ai are with the School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: yiernanh@163.com; hjq099@163.com; xiaobin_792002@163.com).

X. Li is with the School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China, and also with the Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131 USA (e-mail: lixia@mail.sysu.edu.cn; lixia@graduate.hku.hk).

L. Liu is with the Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131 USA (e-mail: lin.liu@uc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2008.2001754

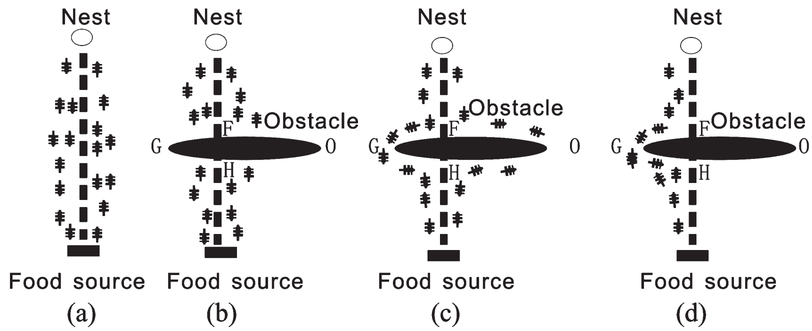



Fig. 1. Route choice behaviors of ants in seeking foods.

and simpler rules than some decision tree methods. Compared with traditional statistical methods, Ant-Miner has a number of advantages [23], [26]. First, Ant-Miner is distribution free, which does not require training data to follow a normal distribution. Second, Ant-Miner is a rule induction algorithm, which is more explicit and comprehensible than mathematical equations. Finally, Ant-Miner requires minimum understanding of the problem domain.

Ant-Miner is different from decision tree approaches such as See 5.0. The entropy measure is a local heuristic measure in See 5.0 [27], which considers only one attribute at a time, and so it is sensitive to attribute correlation problems. Whereas in Ant-Miner, pheromone updating tends to cope better with attribute correlation, since pheromone updating is directly based on the performance of the rule as a whole [25]. Thus, Ant-Miner should have great potential in improving remote-sensing classification because of these advantages. In this paper, an Ant-Miner program for discovering classification rules is developed for the classification of remote-sensing images. It can discover optimized classification rules through simulating the behavior of ants seeking foods. A discretization technique is incorporated in the model to improve the performance of classifications involving a large set of data.

## II. ACO

ACO is based on the ants' behaviors in finding the shortest path when seeking food without the benefit of visual information [17]. Ethologists have discovered that to exchange information about which path to follow, ants communicate by releasing pheromone along their routes. Ants choose a path that has the largest amount of pheromone. Since pheromone decays in time, a shorter route will have a higher concentration of pheromone than a longer route. The attraction of more ants to a shorter route further increases the concentration of pheromone along the shorter route. This way, ants are capable of finding the shortest route from their nests to food sources without using visual cues. This process can be described as a loop of positive feedback, in which the probability of an ant choosing a path is proportional to the number of ants that have already passed by that path [17].

This positive feedback mechanism makes ACO self-adaptive. The process of seeking food by an ant colony is illustrated in Fig. 1. If there are no obstacles between ant nests and food sources, the shortest route is in a straight line [Fig. 1(a)]. If an obstacle cuts off the straight path at location F [Fig. 1(b)], ants

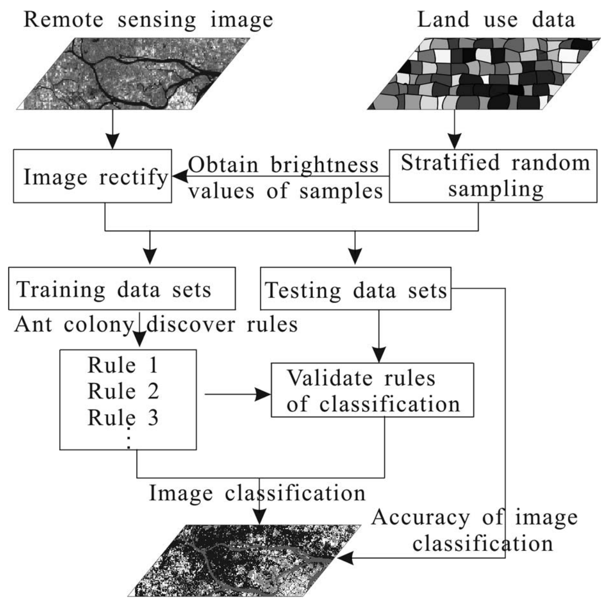



Fig. 2. Flow chart of Ant-Miner-based classification of remote-sensing images.

select various routes with an identical probability and deposit pheromone on the selected routes. Since the route F–G–H is shorter than F–O–H, the ants selecting the route F–G–H reach the food source sooner than those selecting the route F–O–H. As a result, H–G–F has a higher concentration of pheromone than on H–O–F, and it will, therefore, attract more ants [Fig. 1(c)]. At the final stage, all the ants will choose the route H–G–F because the pheromone on the longer route gradually disappears [Fig. 1(d)].

## III. ANT-MINER FOR REMOTE-SENSING CLASSIFICATION

This paper modifies and extends ACO to rule induction for classifying remote-sensing images (Fig. 2). In ACO-based rule induction, the mapping from attributes to classes is analogous to the route search by an ant colony. An attribute node corresponds to a brightness value of remote-sensing images. An attribute node can only be selected once and must be associated with a class node. As shown in Fig. 3, each route corresponds to a classification rule, and discovering a classification rule can be regarded as searching for an optimal route. A rule can randomly be generated at the start. The rule can be represented as

$$\text{IF } \langle \text{term}_1 \text{ AND term}_2 \text{ AND} \ldots \rangle \text{ THEN } \langle \text{class} \rangle \quad (1)$$

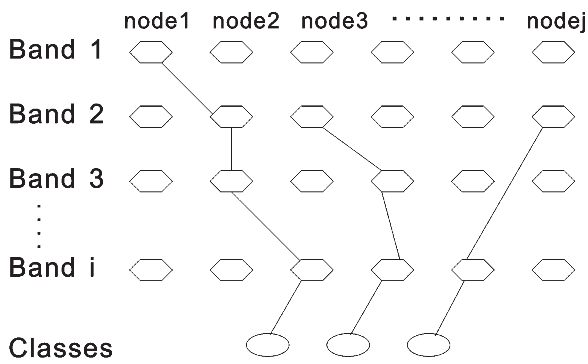



Fig. 3. Route corresponding to classification rule derived from Ant-Miner.

where $\langle \text{term}_1 \text{ AND term}_2 \text{ AND} \ldots \rangle$ are conditions that contain the terms using the logical operator AND. Each term can be expressed as a triplet $\langle \text{band, operator, brightness value} \rangle$. $\langle \text{class} \rangle$ is the prediction of the class.

It should be noted that the original values of remote-sensing images must be divided into a finite number of intervals by using a discretization technique for facilitating the route search. For example, if the brightness value of band $B_i$ has a range from 0 to 255, the discretization process divides the range into discrete intervals like (0–13), (14–25), (26–41), and so on. This can reduce the number of possible rules and help improve the efficiency of the ACO algorithm. The following sections provide a detailed procedure for applying ACO to remote-sensing classification.

### A. Discretization of Remote-Sensing Data

The existing rule induction algorithms, such as decision trees and rough sets, usually deal with discrete data. Continuous attributes must be discretized into a finite number of levels or discrete values [28]. For example, See 5.0 constructs the classification trees from discrete values based on the "information gain" calculated by the entropy [29]. CART applies the Gini criterion to discretize continuous attributes [30]. SIPINA takes advantage of the Fusinter criterion, which is based on the measurement of uncertainty [31]. Many applications involve the use of continuous attributes, which cannot be directly processed by these algorithms. Discretization is an effective technique in dealing with continuous attributes for rule generating [32]. This procedure increases the speed and accuracy of machine learning [33]. In general, results obtained through decision trees or induction rules using discretized data are usually more efficient and accurate than those using continuous values [34].

Although the brightness values of remote-sensing data are discrete integers (0–255), there are still too many unique values for each band. This will influence the efficiency of the Ant-Miner algorithm and the quality of rules. A discretization technique is applied in this paper to divide the original brightness values into a smaller number of intervals. Selecting the proper methods for this transformation is very important because it determines the overall quality of generating rules. In this paper, an entropy method is adopted to measure the importance of breakpoints for the discretization of brightness values [35].

For a convenient expression, a decision table is defined as a table of information comprised of a four-element set $(U, R, V, f)$, among which $U$ refers to a set of objects, i.e., domain; $R = C \cup D$, where $C$ and $D$ refer to a condition attribute set and a decision attribute set, respectively. In this paper, $C$ and $D$ refer to the bands of satellite images and the land use types from the classification, respectively. $V$ represents the value range of each band, and $f$ is the information function.

If $X \subseteq U$ is considered as a training subset consisting of $|X|$ samples, among which $k_j$ samples are noted with decision attribute $j (j = 1, 2, \ldots, r)$, then the information entropy for the training set is [35], [36]

$$H(X) = -\sum_{j=1}^{r} P_j \log_2 P_j, \qquad P_j = \frac{k_j}{|X|} \tag{2}$$

where a smaller value of the entropy indicates that the set $X$ is determined by several dominant values of decision attributes and a smaller degree of disorder. The total number of class is denoted by $N$. $c_i^a$ is the $i$th breakpoint selected from band $a$. Samples of decision attribute $j (j = 1, 2, \ldots, r)$ belonging to the set $X$ can be divided into two types, i.e., those with an attribute value smaller than $c_i^a$ and those with an attribute value equal and greater than $c_i^a$.

Consequently, the set $X$ is classified as $X_l$ and $X_r$, whose information entropies are calculated as follows [32], [33]:

$$H(X_l) = -\sum_{j=1}^{r} p_j \log_2 p_j, \qquad p_j = \frac{l_j^X\left(c_i^a\right)}{l^X\left(c_i^a\right)} \tag{3}$$

$$H(X_r) = -\sum_{j=1}^{r} q_j \log_2 q_j, \qquad q_j = \frac{r_j^X\left(c_i^a\right)}{r^X\left(c_i^a\right)} \tag{4}$$

where

$$l^X\left(c_i^a\right) = \sum_{j=1}^{r} l_j^X\left(c_i^a\right) \tag{5}$$

$$r^X\left(c_i^a\right) = \sum_{j=1}^{r} r_j^X\left(c_i^a\right) \tag{6}$$

and where $l_j^X(c_i^a)$ is the total number of cases for class $j$ with an attribute value smaller than $c_i^a$, and $r_j^X(c_i^a)$ is the total number of cases for class $j$ with an attribute value equal and greater than $c_i^a$.

Table I provides a simple example to explain the meanings of $l_j^X(c_i^a)$ and $r_j^X(c_i^a)$ in the situation of ten cases. It is assumed that breakpoint 60 is the second breakpoint of band 4 $(i = 2)$, and the number of classes is 5 $(N = 5)$. The following results can then be derived: $l_1^X(c_i^a) = 2$, $l_2^X(c_i^a) = 2$, $l_3^X(c_i^a) = 1$, $l_4^X(c_i^a) = 0$, $l_5^X(c_i^a) = 1$; $r_1^X(c_i^a) = 0$, $r_2^X(c_i^a) = 0$, $r_3^X(c_i^a) = 0$, $r_4^X(c_i^a) = 4$, $r_5^X(c_i^a) = 0$. According to (5) and (6), $l^X(c_i^a) = 2 + 2 + 1 + 1 + 0 = 6$, and $r^X(c_i^a) = 0 + 0 + 4 + 0 = 4$.

Additionally, the information entropy of breakpoint $c_i^a$ relative to the set $X$ is defined as

$$H^X\left(c_i^a\right) = \frac{|X_l|}{|U|} H(X_l) + \frac{|X_r|}{|U|} H(X_r). \tag{7}$$

Supposing that $L = \{Y_1, Y_2, \ldots, Y_m\}$ is the equivalent sample derived from the division of the set $P$ with breakpoints

TABLE I
CLASS OF TEN CASES WITH THE VALUE IN BAND 4

| Case | case1 | case2 | case3 | case4 | case5 | case6 | case7 | Case8 | Case9 | Case10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Band 4 | 10 | 15 | 18 | 30 | 38 | 40 | 80 | 85 | 150 | 180 |
| Class | 1 | 1 | 2 | 2 | 3 | 5 | 4 | 4 | 4 | 4 |

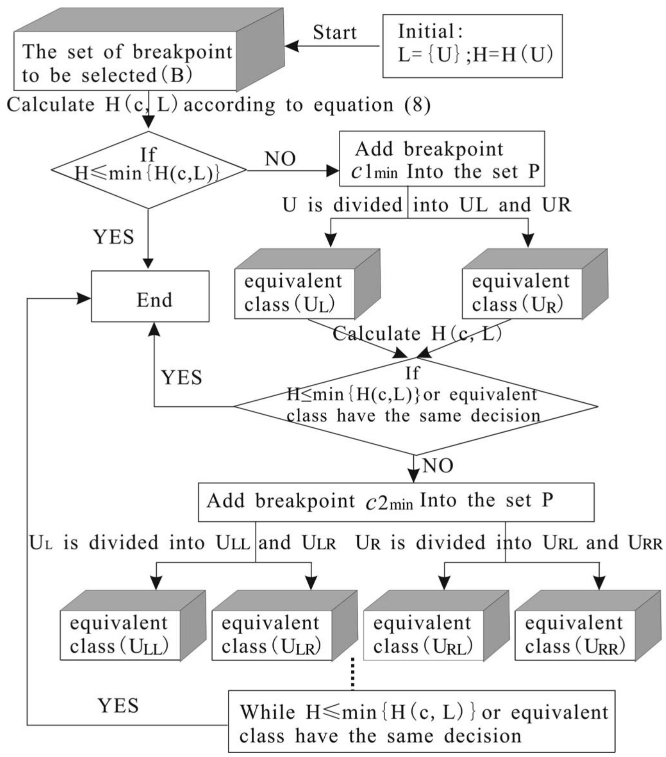



Fig. 4. Flow chart of discretizing attribute values according to information entropy.

selected from the decision table, then the new information entropy after the addition of breakpoint $c \notin P$ becomes

$$H(c, L) = H^{Y_1}(c) + H^{Y_2}(c) + \cdots + H^{Ym}(c) \tag{8}$$

where a smaller $H(c, L)$ indicates that the decision attribute value of the new equivalent subset divided tends to be more monotonous after adding the breakpoint, which will be more important.

If $P$ is defined as the set of breakpoints, $L$ as the set of equivalent samples divided by the breakpoint set $P$, and $B$ as the set of breakpoints to be selected, for each $c \in B$, $band_{\min} < c < band_{\max}$. With $H$ as the information entropy of the decision table, the process of discretizing attribute values according to information entropy can be described as follows [35] (Fig. 4).

1) For each $c \in B$, calculate $H(c, L)$.
2) If $H \le \min H(c, L)$, then end.
3) Select and add breakpoint $c_{\min}$, which can make $H(c, L)$ minimum into the set $P$.
4) For all $X \in L$, if an equivalent $X$ is divided into $X_1$ and $X_2$ with $c_{\min}$, then $X$ can be removed from $L$, whereas the equivalent classes $X_1$ and $X_2$ can be added into $L$.
5) If each equivalent sample among $L$ shows the same decision, terminate the loop. If not, return to step 1.

## *B. Rule Construction Using Ant-Miner*

Ant-Miner has been applied to the discovery of classification rules by using these discretized data. The rules are discovered according to an approach similar to the collective process of seeking foods by ants. Ant-Miner uses a sequential covering algorithm to discover a list of rules that cover all or most of the training samples (pixels) in the training set. At first, the list of rules is empty, then Ant-Miner obtains a set of ordered rules through iteratively finding a "best" rule that covers a subset of the training set. Next, Ant-Miner adds this "best" rule to the discovered rule list and removes the training samples covered by the rule until a stop criterion is reached. The ordered rule is applied to a training sample only if none of the previous rules in the rule set are applicable [37].

Term selection is an important step for Ant-Miner in constructing rules. Theoretically, term selection can completely be random, but this search involves intensive computation. A heuristic function is designed to guide the search so that the computation time can greatly be reduced. The information entropy is used to define this function, in which the heuristic value for each term is proportional to its classification capability [25]. In this paper, a heuristic function based on the statistical attribute of the data (frequency) is designed, in which the heuristic value $\eta_{ij}$ of the condition item $\text{term}_{ij}$ is defined as follows [22]:

$$\eta_{ij} = \frac{\max\left(\sum_n freqT_{ij}^1, \sum_n freqT_{ij}^2, \ldots, \sum_n freqT_{ij}^k\right)}{\sum_n T_{ij}} \tag{9}$$

where $\eta_{ij}$ denotes the density-based heuristic value of $\text{term}_{ij}$, and $\text{term}_{ij}$ is the condition item of the classification rule. $T_{ij}$ refers to the number of training samples fitting to this condition term, and $freqT_{ij}^w$ is the frequency of class $w$ in $T_{ij}$. The record that satisfies the condition part of the rule should be removed after a final rule has been obtained. Therefore, the values for $\max(\sum_n freqT_{ij}^1, \sum_n freqT_{ij}^2, \ldots, \sum_n freqT_{ij}^k)$ and $\sum_n T_{ij}$ are updated after a final rule has been found.

The other two parameters, i.e., the amount of pheromone and the probability of terms to be selected, are also important to the rule construction. When a route is found by an artificial ant, the amount of pheromone for all the nodes in this route will be initialized to the same value as

$$\tau_{ij}(t = 0) = \frac{1}{\sum_{i=1}^{A} b_i} \tag{10}$$

where $\tau_{ij}$ is the amount of pheromone for the condition term ($\text{term}_{ij}$), $A$ is the sum of attributes (excluding the class attributes) in the data bank, and $b_i$ refers to any possible value of attribute $i$.

The roulette wheel selection technique is adopted to decide which term will be included for constructing a path according to the heuristic value $(\eta_{ij})$ and the thickness of pheromone $(\tau_{ij})$. The probability that a term will be added to the current rule is given by

$$P_{ij}(t) = \frac{\tau_{ij}(t) \cdot \eta_{ij}(t)}{\sum_{i=1}^{a} \sum_{j=1}^{b_i} \tau_{ij}(t) \cdot \eta_{ij}(t)}. \quad (11)$$

The probability of a term being selected depends on both frequency and pheromone updating. This makes the rule construction process of Ant-Miner more robust and less prone to get trapped into the local optima in the search space, since the feedback provided by pheromone updating helps to correct some mistakes made by the shortsightedness of the frequency measure. Whereas in decision tree algorithms, the entropy measure is the only heuristic function used during tree building [25]. The selected term will be added to the rule until all attributes are selected to form a complete classification rule. The validity of this rule can be assessed by using the following equation [25]:

$$Q = \left(\frac{\text{TruePos}}{\text{TruePos} + \text{FalseNeg}}\right) \cdot \left(\frac{\text{TrueNeg}}{\text{FalsePos} + \text{TrueNeg}}\right) \quad (12)$$

where TruePos (true positives) is the total number of positive cases correctly predicted by the rule, FalsePos (false positives) is the total number of positive cases wrongly predicted by the rule, TrueNeg (true negatives) is the total number of negative cases correctly predicted by the rule, and FalseNeg (false negatives) is the total number of negative cases wrongly predicted by the rule. The larger the value of $Q$, the higher the quality of the rule.

### C. Rule Pruning

The next step is to prune the discovered rules for improving the classification performance. Rule pruning is a common technique in rule induction [38]. The main goal of rule pruning is to remove irrelevant terms, since a short rule is, in general, more comprehensible by the user than a long rule [25]. Another motivation for rule pruning is to improve the predicative accuracy of rules [25] because some rules make little contribution to the classification and may even reduce the overall accuracy. Furthermore, rule pruning prevents the rules from overfitting the training data. The basic idea of rule pruning is to iteratively remove one term at a time from a rule while this process improves the quality of the rule [25]. The removal process is repeated until only one term remains in the rule antecedent or the rule quality no longer improves. The rule quality is defined by (12).

### D. Pheromone Updating

The pheromones of all terms are initialized with an equal value according to (10). Once a rule becomes acceptable, the amount of pheromone for all terms is updated. The pheromone of terms that occur in the constructed rule increases according to the quality of classification rule. In contrast, the amount of pheromone associated with the terms not included in the constructed rule decreases. The rate of drop is determined by the evaporation coefficient $\rho$. The amount of pheromone of each term is updated according to the following equation:

$$\tau_{ij}(t+1) = (1-\rho) \cdot \tau_{ij}(t) + \left(\frac{Q}{1+Q}\right) \cdot \tau_{ij}(t) \quad (13)$$

where $\rho$ is the pheromone evaporation coefficient, and $Q$ is the quality of a classification rule.

After the amount of pheromone for all terms is updated, the next artificial ant starts a new round of search. The search converges when the majority of ants locate a route for efficient food seeking. The iteration continues until all ants complete their search. During each iteration, these ants may construct many rules, of which only the best rule is preserved, and the others are discarded. This process is repeated until the number of remaining training classes is less than the predefined number.

The detailed procedure of discovering the rules for remote-sensing classification is as follows.

1) Discretizing the brightness value for each band of remote-sensing images;
2) Starting from an empty route for an ant, adding nodes to this route to find a complete route according to the amount of pheromone at each node.
3) When an ant passes a route, it releases an amount of pheromone at the nodes according to the travel time (cost). The amount of pheromone will affect the probability of selecting this route by other ants.
4) Pruning the rules (routes) generated by the collective behavior of ants.
5) During each iteration, updating the amount of pheromone for all terms. This provides feedback for the next round of search.
6) Go to step 2 until all ants have been examined in route search.
7) Choose the best rule among all the rules constructed by these ants.
8) Remove the set of training samples correctly covered by the final rule discovered by step 7.
9) Go to step 2 until the number of remaining training samples becomes less than a threshold value.

The detailed computation algorithm is listed as follows.

**ALGORITHM**: A High-level description of the Ant-Miner for classification of remote-sensing images

The original trainingSet
Discretization of the original TrainingSet
DiscoveredRuleList $=$ []/ $*$ rule list is initialized with an emptylist $*$ /
WHILE (TrainingSet $>$ Max_uncovered_training samples)
Initialize all nodes with the same amount of pheromone
    Calculation the $\eta_{ij}$ of the training data for all nodes
  $i = 1$   / $*$ ant index $*$ /
  WHILE ( $i <$ No_of_ants and $m <$ No_rules_converg)
    FOR $j = 1$ to No_of_attributes
      Select a node of the attribute

Fig. 5. TM image (5, 4, 3) in the study area of Guangzhou.

NEXT $j$
Obtaining $\mathrm{Rule_i}$
Rules pruning
IF ($\mathrm{Rule_i}$ is equal to $\mathrm{Rule_{i-1}}$)
THEN $m = m + 1$
ELSE $m = 1$
END IF
Pheromone update
$i = i + 1$
LOOP
Select the best rule $\mathrm{R_{best}}$ among all rules and add it to DiscoveredRuleList;
TrainingSet = TrainingSet − {set of training samples covered by $\mathrm{R_{best}}$}
LOOP

## IV. Model Implementation and Results

A satellite Landsat TM image of the Guangzhou area acquired on July 18, 2005 is used for the experiment of classification using this ACO method. The study area consists of 1706 × 1549 pixels with a ground resolution of 30 m (Fig. 5). This area is dominated by the following six land use types: residential, forest, water, orchard (banana and sugarcanes), cropland, and developing land. Selection of proper training samples is a key step for ant colony learning and is directly related to the quality of the discovered rules. Based on field investigation and land use maps, a total of 2150 samples (pixels) are acquired by using a hierarchically random sampling method [39]. The sample data set is further divided into two groups, i.e., 1000 as the training data set and 1150 as the test data set.

The ACO classification model involves a two-step procedure, i.e., discovering classification rules from the training data and obtaining land use types for the remote-sensing image. Classification rules are discovered from the training data through the Ant-Miner program, which is developed using the Visual Basic 6.0 language. Land use types are obtained by applying these rules discovered by the Ant-Miner to the classification of remote-sensing images.

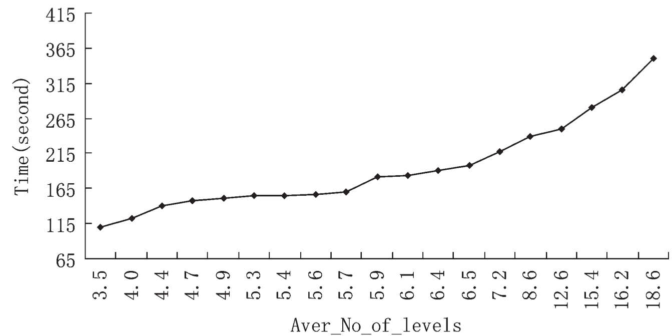



Fig. 6. Influence of Aver_No_of_levels on the efficiencies of Ant-Miner.

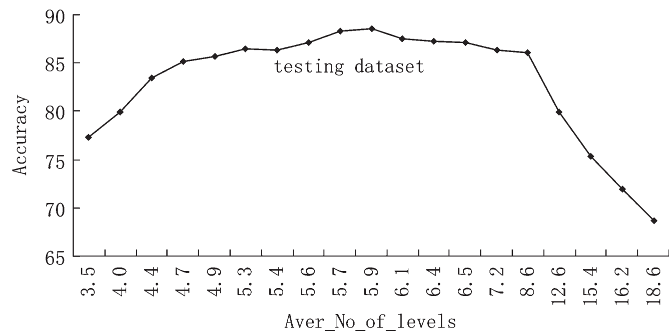



Fig. 7. Influence of Aver_No_of_levels on the accuracy of Ant-Miner.

In the training data set, brightness values are the attribute nodes of routes, and the known land use types are the class nodes. These bands have to be discretized before Ant-Miner is applied to the classification of remote-sensing data. Experiments carried out in this paper reveal that this discretization procedure actually decreases the computation complexity and improves the accuracy of classification. Classification efficiencies decrease with the increase of the average number of intervals in the discretized data (Aver_No_of_levels) (Fig. 6).

As shown in Fig. 7, the classification accuracy of Ant-Miner improves when the average number of intervals increases from 3.5 to 5.9. The further accuracy improvement is not obvious when Aver_No_of_levels increases from 5.9 to 8.6. Actually, the accuracy decreases after the average number of intervals becomes greater than 8.6. Therefore, a higher classification accuracy can be obtained if these original brightness values are properly discretized. It is much better to use these discretized brightness values instead of the original brightness values (0–255). In this paper, the original levels are reduced to 8.6 levels (on average) after the discretization. The reduced levels are very close to those of the C4.5 method. In Parpinelli's research [25], the discretization was performed by the C4.5-Disc discretization method. It is found that the original levels are reduced to 9.7 levels (on average) if this C4.5-Disc discretization method is used to discretize the remote-sensing data of this study area.

This Ant-Miner program requires the specification of the following parameters.

1) No_of_ants (Number of ants): This is the maximum number of candidate rules constructed during an iteration.
2) Min_training samples_per_rule (Minimum number of training samples per rule): This is the minimum number

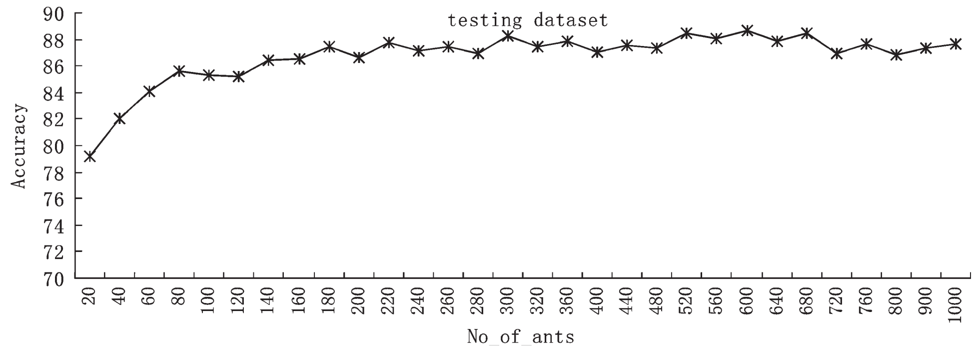



Fig. 8. Influence of No_of_ants on the performance of Ant-Miner.

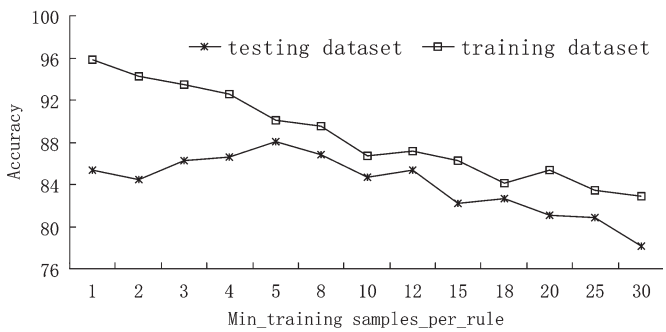



Fig. 9. Influence of Min_training samples_per_rule on the performance of Ant-Miner.

of training samples each rule must at least cover, which helps avoid overfitting the training data.

3) Max_uncovered_training samples (Maximum number of uncovered training samples in the training set): The process of discovering rules is iteratively performed until the number of uncovered training samples is smaller than this threshold.
4) Max_iterations (Maximum number of iterations): The program stops when the number of iterations is larger than this threshold.

The default parameter settings for the Ant-Miner are the following: No_of_ants $= 180$; Min_training samples_per_rule $= 5$; Max_uncovered_training samples $= 20$; and Max_iterations $= 200$. The experiment indicates, among these four parameters, that the number of ants (No_of_ants) and the minimum number of training sample per rule (Min_training samples_per_rule) are the two most sensitive factors in determining classification results. The sensitivity of these two parameters is shown in Figs. 8 and 9. Classification results improve with the increase of No_of_ants. This improvement stabilizes after No_of_ants reaches 180 (Fig. 8). As shown in Fig. 9, the larger is Min_training samples_per_rule, the lower the accuracy of the training set becomes. However, there are slight differences between the test set and the training set. The relationship changes when Min_training samples_per_rule is smaller than 5. When Min_training samples_per_rule $= 5$, Ant-Miner has the highest predictive accuracy.

A total of 44 classification rules are generated by the Ant-Miner, which takes 3 min to complete the rule induction by using the training data. It takes much longer to discover rules than See 5.0. A selected set of the classification rules is listed in Table II. Fig. 10(a) shows the classified remote-sensing image based on this proposed method.

TABLE II
PARTS OF THE CLASSIFICATION RULES BY USING ANT-MINER

| |
|---|
| Rule 1: |
| IF |
| $54<B_4<129$ and $132<B_5<203$ and $45<B_7<62$ |
| Then |
| class=Residential (credibility =1.0) |
| Rule 2: |
| IF |
| $106<B_1<113$ and $60<B_3<97$ and $54<B_4<130$ and $41<B_5<132$ |
| Then |
| class= Orchard (credibility =0.93) |
| Rule 3: |
| IF |
| $B_3<60$ and $129<B_4<160$ |
| Then |
| class=Forest (credibility =1.0) |
| Rule 4: |
| IF |
| $B_4<54$ and $B_7<14$ |
| Then |
| class=Water (credibility =1.0) |
| …… |

The results from this ACO method are compared with those from the See 5.0 decision tree method. The decision tree method automatically discovers classification rules by using machine learning techniques. It uses the "information gain ratio" to determine the splits at each internal node of the decision tree [27]. The comparison of these methods is carried out through three criteria, namely, the simplicity of the discovered rule list, the overall classification accuracy, and the Kappa coefficient. The simplicity of rules is measured by the number of discovered rules and the average number of terms per rule [25]. In comparison, the same training data (1000 samples) are used for the classification, and the same test data (1150 samples) were used for validation. The results comparing the simplicity of the rule lists discovered by ACO and See 5.0 are reported in Table III. The ACO method discovers a compact rule list with 44 rules and 2.54 terms per rule, whereas See 5.0 discovered a rule list with 61 rules and 2.73 terms per rule, respectively, which indicates that the ACO-discovered rules are simpler than the rules discovered by See 5.0.

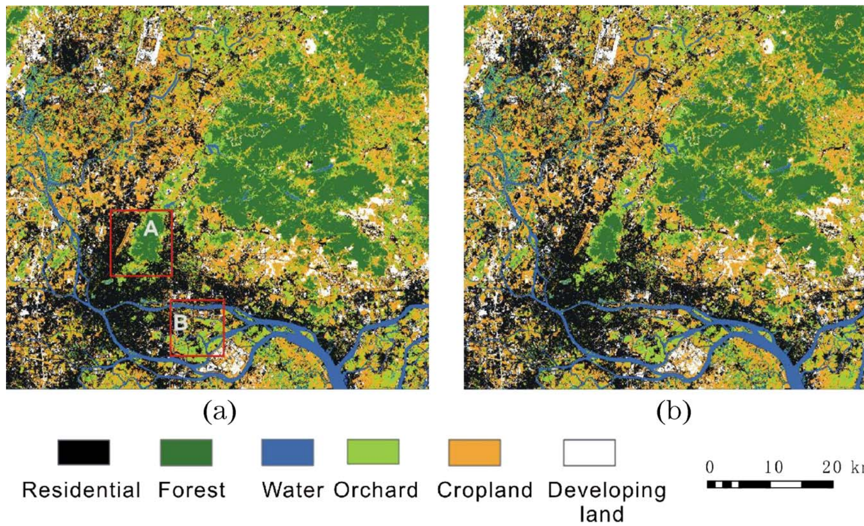



Fig. 10. TM image and land use classification in the study area of Guangzhou.

TABLE III
SIMPLICITY OF RULE LISTS DISCOVERED BY ACO AND SEE 5.0

| The number of rules | | The average number of terms per rule | |
|---|---|---|---|
| ACO | See 5.0 | ACO | See 5.0 |
| 44 | 61 | 2.54 | 2.73 |

The classification result of the decision tree method using the See 5.0 system is shown in Fig. 10(b). The comparison between Fig. 10(a) and (b) indicates that the ACO method is better than the See 5.0 decision tree method. An enlarged part of the study area [A and B in Fig. 10(a)] is shown in Fig. 11, where Fig. 11(a) is the original remote-sensing image. Area $A$ in Fig. 11 is actually forest land, but a part of area $A$ is incorrectly classified as orchard by the decision tree method. Some orchards were incorrectly classified as cropland by the decision tree method in Fig. 11(c). However, these land use types have been correctly classified by the ACO method [see $A$ and $B$ in Fig. 11(b)].

As shown in Tables III and IV, the total accuracy is 88.6% by using the Ant-Miner method. In contrast, a lower total accuracy (85.4%) is obtained by using the See 5.0 system. However, it is well recognized that although widely used, the total accuracy is not by itself a perfect measure of classifier performance [40]. A more reasonable measure is the Kappa coefficient, since it takes account of the chance occurrence of correct classifications. As shown in Tables IV and V, the kappa coefficients of Ant-Miner and See 5.0 are 0.861 and 0.822, respectively. Taking into account the rule list simplicity, the overall classification accuracy, and the kappa coefficient, the results of our experiments indicate that the Ant-Miner method yields better accuracy and simpler rule sets than the See 5.0 system.

## V. CONCLUSION

Intelligent methods can improve the performance of remote-sensing classification. Traditional methods have some limitations in constructing proper classifiers for remote-sensing classification if the study area is complex. For example, statistical classification methods require the data to follow a normal distribution, but this assumption may not be valid. This paper has presented a new method to classify remote-sensing data by using ACO. A technique of discretization has also been proposed for the efficient retrieval of classification rules. Furthermore, an Ant-Miner program has been developed to derive classification rules for remote-sensing images. The derived rule sets are more explicit and comprehensible than the mathematical equations of using statistical methods.

ACO is, in fact, a complex multiagent system in which agents with simple intelligence can complete complex tasks through cooperation. It can deal with difficult classification problems. Ant intelligence is based on pheromone updating for optimization. The classification rules derived by ant intelligence can easily be represented without using complex equations. Compared with the See 5.0 decision tree method, the Ant-Miner tends to cope better with attribute correlation, since its pheromone updating is directly based on the global (as opposed to local) performance of the rule. As a result, the Ant-Miner is capable of providing better classification results.

This method has been applied to the classification of remote-sensing images in Guangzhou, China. The comparison of classification results is carried out between the ACO method and the See 5.0 decision tree method. The overall accuracy of the ACO method is 88.6% with a Kappa coefficient of 0.861. The decision tree method has an accuracy of 85.4% and a Kappa coefficient of 0.822. Furthermore, the rule sets discovered by ACO are more succinct than those by See 5.0. Therefore, this ACO method is more effective for the classification of remote-sensing images than the See 5.0 decision tree method.

In this paper, Ant-Miner has been successfully applied to classification of remote sensing. However, there are still some limitations in using this method to discover rules. First, the rules generated by the Ant-Miner have a large number of boxes in the feature space. This is because the condition part of the rule only contains the term using the logical operator AND.

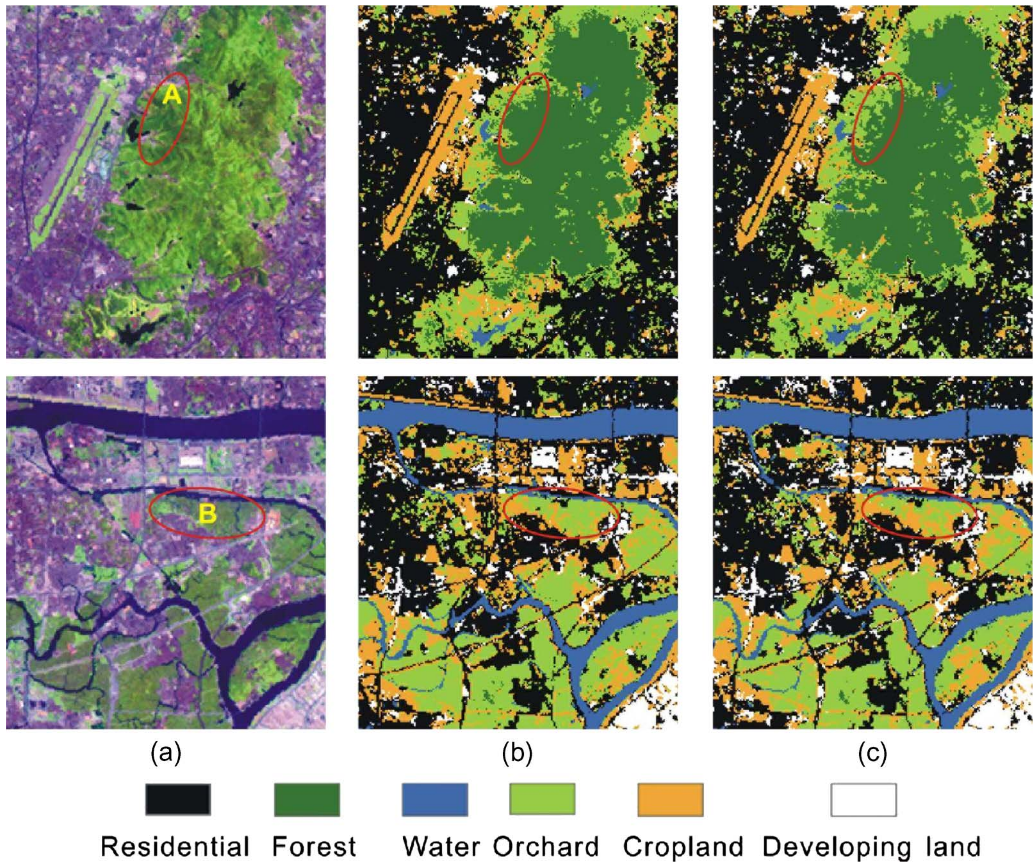



Fig. 11. Land use classification in the local enlargement area of Guangzhou.

TABLE IV
ACCURACY ASSESSMENT ON ACO-BASED CLASSIFICATION IN THE STUDY AREA OF GUANGZHOU

| class | residential | forest | water | orchard | cropland | Developing land | Total | Producer's accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| residential | 266 | 0 | 4 | 3 | 14 | 7 | 294 | 90.5 |
| forest | 1 | 165 | 0 | 5 | 3 | 1 | 175 | 94.3 |
| water | 2 | 0 | 124 | 2 | 2 | 0 | 130 | 95.4 |
| orchard | 8 | 17 | 2 | 153 | 16 | 2 | 198 | 77.3 |
| cropland | 15 | 4 | 1 | 14 | 176 | 3 | 213 | 82.6 |
| Developing land | 3 | 0 | 0 | 1 | 1 | 135 | 140 | 98.5 |
| Total | 295 | 186 | 131 | 178 | 212 | 148 | 1150 | |
| User's accuracy (%) | 90.17 | 88.7 | 94.7 | 84.8 | 83.0 | 91.2 | | |
| | | Total accuracy = 88.6% | | | Kappa coefficient = 0.861 | | | |

TABLE V
ACCURACY ASSESSMENT FOR THE SEE 5.0 DECISION TREE METHOD IN THE STUDY AREA OF GUANGZHOU

| class | residential | forest | water | orchard | cropland | Developing land | Total | Producer's accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| residential | 259 | 0 | 4 | 3 | 23 | 9 | 298 | 86.9 |
| forest | 2 | 156 | 2 | 6 | 4 | 1 | 171 | 91.2 |
| water | 2 | 0 | 122 | 4 | 3 | 0 | 131 | 93.1 |
| orchard | 8 | 24 | 1 | 147 | 17 | 2 | 199 | 73.9 |
| cropland | 19 | 6 | 2 | 17 | 164 | 2 | 210 | 78.1 |
| Developing land | 5 | 0 | 0 | 1 | 1 | 134 | 141 | 95.0 |
| Total | 295 | 186 | 131 | 178 | 212 | 148 | 1150 | |
| User's accuracy (%) | 87.8 | 83.9 | 93.1 | 82.5 | 77.4 | 90.5 | | |
| | | Total accuracy = 85.4% | | | Kappa coefficient = 0.822 | | | |

In future research, the logical operator OR should be added in the conditions of the rules. In fact, XOR is a difficult problem in rule induction algorithms. Even the See 5.0 method only uses the logical operator AND in the conditions of the rules [27]. Second, Ant-Miner uses a sequential covering algorithm to discover rules, so the rules are ordered. This makes it difficult to interpret the rules at the end of the list, since the meaning of a rule in the list is dependent on all the previous rules. Finally, this Ant-Miner takes a much longer time to discover rules than the See 5.0 method.

## REFERENCES


[1] J. A. Richards and X. P. Jia, *Remote Sensing Digital Image Analysis, An Introduction*. Berlin, Germany: Springer-Verlag, 1999. 3rd revised and enlarged edition.

[2] G. G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 433–440, Mar. 2005.

[3] B. Turner, W. B. Meyer, and D. L. Skole, "Global land-use/land-cover change: Toward an integrated study," *Ambio*, vol. 23, no. 1, pp. 91–95, 1994.

[4] G. G. Wilkinson, "A review of current issues in the integration of GIS and remote sensing data," *Int. J. Geogr. Inf. Sci.*, vol. 10, no. 1, pp. 85–101, Jan. 1996.

[5] A. H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sens. Environ.*, vol. 10, no. 2, pp. 135–163, Sep. 1980.

[6] S. Di Zenzo, R. Bernstein, S. D. Degloria, and H. G. Kolsky, "Gaussian maximum likelihood and contextual classification algorithms for multicrop classification," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-25, no. 6, pp. 805–814, Nov. 1987.

[7] J. Ton, J. Sticklen, and A. Jain, "Knowledge-based segmentation of Landsat images," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 2, pp. 222–232, Mar. 1991.

[8] K. Muthu and M. Petrou, "Landslide-hazard mapping using an expert system and a GIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 522–531, Feb. 2007.

[9] R. Tonjes, S. Growe, J. Buckner, and C. E. Liedtke, "Knowledge-based interpretation of remote sensing images using semantic nets," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 7, pp. 811–821, 1999.

[10] L. Bruzzone and D. F. Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote sensing images," *IEEE Trans. Geosci Remote. Sens.*, vol. 37, no. 2, pp. 1179–1184, Mar. 1999.

[11] L. Jimenez, A. Morales-Morell, and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1360–1366, May 1999.

[12] F. Del Frate, F. Pacifici, G. Schiavon, and C. Solimini, "Use of neural networks for automatic classification from high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 800–809, Apr. 2007.

[13] D. Stathakis and A. Vasilakos, "Comparison of computational intelligence based classification techniques for remotely sensed optical image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2305–2318, Aug. 2006.

[14] A. Colorni, M. Dorigo, V. Maniezzo *et al.* "Distributed optimization by ant colonies," in *Proc. 1st Eur. Conf. Artif. Life*, 1991, pp. 134–142.

[15] R. Eberhart, Y. Shi, and J. Kennedy, *Swarm Intelligence*. San Francisco, CA: Morgan Kaufmann, 2001.

[16] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997.

[17] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. dissertation, Dept. Electron., Politecnico di Milano, Milan, Italy, 1992.

[18] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996.

[19] C. Blum and K. Socha, "Training feed-forward neural networks with ant colony optimization: An application to pattern classification," in *Proc. 5th Int. Conf. Hybrid Intell. Syst.*, 2005, pp. 233–238.

[20] K. M. Sim and W. H. Sun, "Multiple ant-colony optimization for network routing," in *Proc. 1st Int. Symp. Cyber Worlds*, 2002, pp. 277–281.

[21] E. Lumber and B. Faieta, "Diversity and adaptation in populations of clustering ants," in *Proc. 3rd Int. Conf. Simul. Adaptive Behavior: From Animals to Animates*, 1994, pp. 501–508.

[22] H. B. Duan, *Ant Colony Algorithms: Theory and Applications*. Beijing, China: Science Press, 2005.

[23] B. Alatas and E. Akin, "FCACO: Fuzzy classification rules mining algorithm with ant colony optimization," in *Proc. Advances Natural Comput.*, 2005, vol. 3612, pp. 789–797.

[24] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "An ant colony algorithm for classification rule discovery," in *Data Mining: A Heuristic Approach*, H. A. Abbass, R. A. Sarker, and C. S. Newton, Eds. London, U.K.: Idea Group Publishing, 2002.

[25] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, " Data mining with an ant colony optimization algorithm," *IEEE Trans. Evol. Comput.*, vol. 6, no. 4, pp. 321–332, Aug. 2002.

[26] A. Chan and A. Freitas, "A new classification-rule pruning procedure for an ant colony algorithm," in *Proc. Artif. Evolution*, 2006, vol. 3871, pp. 25–36.

[27] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Boston, MA: Kluwer, 1993, p. 302.

[28] R. P. Li and Z. O. Wang, "An entropy-based discretization method for classification rules with inconsistency checking," in *Proc. 1st Int. Conf. Mach. Learn. Cybern.*, 2002, pp. 243–246.

[29] M. Boulle, "Khiops: A statistical discretization method of continuous attributes," *Mach. Learn.*, vol. 55, no. 1, pp. 53–69, Apr. 2004.

[30] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[31] D. A. Zighed, S. Rabaseda, and R. Rakotomalala, "Fusinter: A method for discretization of continuous attributes for supervised learning," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 33, pp. 307–326, 1998.

[32] C. T. Su and J. H. Hsu, "An extended Chi2 algorithm for discretization of real value attributes," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 437–441, Mar. 2005.

[33] X. Liu and H. Wang, "A discretization algorithm based on a heterogeneity criterion," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 9, pp. 1166–1173, Sep. 2005.

[34] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: An enabling technique," *Data Mining Knowl. Discovery*, vol. 6, no. 4, pp. 393–423, Oct. 2002.

[35] H. Xie, H. Z. Cheng, and D. X. Niu, "Discretization of continuous attributes in rough set theory based on information entropy," *Chin. J. Comput.*, vol. 28, no. 9, pp. 1570–1574, 2005.

[36] H. Theil, *Economics and Information Theory*. Amsterdam, The Netherlands: North Holland, 1967, p. 488.

[37] W. Jiang, Y. Xu, and Y. Xu, "A novel data mining method based on ant colony algorithm," in *Proc. Advanced* Data *Mining Appl.*, 2005, vol. 3584, pp. 284–291.

[38] L. A. Breslow and D. W. Aha, "Simplifying decision trees: A survey," *Knowl. Eng. Rev.*, vol. 12, no. 1, pp. 1–40, Jan. 1997.

[39] C. S. Wu and A. T. Murray, "Estimating impervious surface distribution by spectral mixture analysis," *Remote Sens. Environ.*, vol. 84, no. 4, pp. 493–505, Apr. 2003.

[40] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, Jul. 1991.


**Xiaoping Liu** received the B.S. degree in geography and the Ph.D. degree in remote sensing and geographical information sciences (GIS) from Sun Yat-Sen University, Guangzhou, China, in 2002 and 2008, respectively.

He is currently an Associate Professor with the School of Geography and Planning, Sun Yat-Sen University. His current research interests include image processing, artificial intelligence, and geographical simulation.

**Xia Li** received the B.S. and M.S. degrees from Peking University, Beijing, China, and the Ph.D. degree in GIS from University of Hongkong, Hongkong.

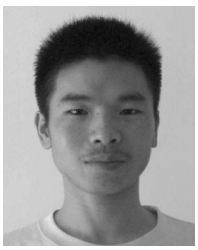
He is a Professor and the Director of the Centre for Remote Sensing and Geographical Information Sciences, School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China. He is also a Guest Professor with the Department of Geography, University of Cincinnati, Cincinnati, OH. He has about 170 articles, of which many appeared in top international journals. He is on the editorial board of the international journal *Geojournal* and is the Associated Editor of a Chinese journal *Tropical Geography*.

Dr. Li received the ChangJiang scholarship and the award of the Distinguished Youth Fund of NSFC in 2005.

**Lin Liu** received the B.S. and M.S. degrees from Peking University, Beijing, China, and the Ph.D. degree in geography from Ohio State University, Columbus.

He is a Professor and the Head of geography with the University of Cincinnati, Cincinnati, OH. His main area of expertise is geographical information science (GIS) and its applications. He became interested in urban simulation in year 2000.

Mr. Liu is a former President of the International Association of the Chinese Professionals in GIS.



**Jinqiang He** received the B.S. degree in geographical information sciences (GIS) in 2006 from Sun Yat-Sen University, Guangzhou, China, where he is currently working toward the Ph.D. degree.

His main research areas include image processing and swarm intelligence.

**Bin Ai** received the M.S. degree in urban geography and GIS from East China Normal University, Shanghai, China, in 2005. She is currently working toward the Ph.D. degree in SAR interferometry at Sun Yat-Sen University, Guangzhou, China.

Her current research interests include temporal radar interferometry and target detection.