A Novel ACO Based Land Cover Classification Approach Using Optical and SAR Data

Qin Dai, Shibin Liu, Jin Yang, and Zhaoming Zhang

Abstract—The land cover classification methods based on statistical theory using remote sensing data have great achievements for the last several decades, but they have exposed some weaknesses in dealing with multi-source and multi-dimensional data. Ant colony optimization (ACO), as an excellent intelligent algorithm, has many potential advantages in the field of remote sensing data processing. In this paper, an intelligent method is developed for classifying the land cover types combining Landsat TM data with Envisat ASAR data on the basis of ACO algorithm. For identifying the classification precision of land cover with ACO algorithm, we respectively compare it with the results by Maximum Likelihood Classification (MLC) and decision tree C4.5. The comparing results show that ACO algorithm can well take advantages of optical and radar data to improve classification precision and select the best features subsets to construct simpler rules.

Index Terms—Ant colony optimizations (ACO), combination of Landsat TM and envisat ASAR data, land cover classification.

I. INTRODUCTION

Land cover is a fundamental parameter that impacts on and links many parts of human and physical environments [1]. Remote sensing image classification has become the main source for land cover maps, and the classification results are the basis for many environmental and socioeconomic applications [2]. For several decades the remote sensing data classification technology has gained great achievements. But dealing with more multi-source and multi-dimensional data, the conventional methods namely parametric methods that assume a normally distributed dataset exists from training samples have exposed some weaknesses. So in recent years, many artificial intelligence techniques were applied to remote sensing data classification field, the purpose of which is to reduce the undesired limitations of the conventional classification methods.

Remote sensing image classification is a very complex process because of many factors, such as the complexity of the landscape in a study area, feature extraction, image pre-processing, selected remote sensing data, and etc. Many researchers have made great efforts in developing advanced techniques for improving classification accuracy [2]. As to the complexity of terrain and remote sensing process, the phenomenon of same terrain shows various spectral features and same spectral feature shows different terrain feature exist in the remote sensing images, which result in complex classification. To perform the problem using multi-feature can make up for only using of single feature classification and improve the precision of classification. Using of multi-feature extracted by optical sensors and Synthetic Aperture Radar (SAR) sensors is a very efficient way. Because data from different sensors can contain distinctive features, using different sensor data simultaneously can take advantage of multiple features of land cover, therefore improve the classification precision [2], [3]. In this paper, the multiple features extracted by optical remote sensing data and SAR data were combined to land cover classification and got a good performance.

As one of the main algorithms in swarm intelligence, Ant Colony Optimization (ACO) takes inspiration from the coordinated behaviour of ant swarms. The real ants have the ability to find the shortest path from their nests to the food source through their communication, and without any visual trace [4]-[6]. ACO was firstly applied on travelling salesman problem which is defined as finding the shortest or nearly shortest path connecting a number of locations, such as cities visited by a travelling salesman on traveller’s sales route. Nowadays, it is used to solve many optimization problems. The ACO algorithm has self-organization, cooperation, communication and other intelligent merits, without prior distributions for variance parameters, can combine multi-source and multi-feature data, so it has great potential advantages in the remote sensing data processing area. Recently although it has been started in remote sensing image classification, it is still a new research topic in the field of remote sensing data processing and has some problems should be resolved. While current Ant-Miner classification algorithm using in the research field can’t cope with continuous attributes directly for remote sensing data classification, always requiring a discretisation reprocessing step before classification [7]. This paper introduces the theory and algorithm process of using ACO algorithm for land cover classification directly based on the continuous DN value, taking full advantage of all continuous attributes’ information.

In this paper, we start by briefly reviewing the ACO algorithm and describe the process aspects of using ACO algorithm to classification rule mining in Section II. In Section III, the experiment of land cover classification using the combination of Landsat TM data and Envisat ASAR data with ACO based algorithm is presented. In Section IV, Conclusions and future work are discussed.
II. METHODS

A. Ant Colony Optimization

As a metaheuristic algorithm, Ant Colony Optimization (ACO) is one of the most successful strands of swarm intelligence, which was introduced by Marco Dorigo and his colleagues in the early 90s inspired by the foraging behaviour of ant colonies [4]-[6]. An individual ant can only do simple tasks on its own, while with the ant colony’s cooperation, they can do complex tasks. In the ACO each ant deposits a chemical on the ground called a pheromone while walking, which guides the future ants, by exploiting pheromone trails ants can find shortest paths between their nest and food sources. As an efficient optimization approach, ACO algorithm has been applied in many domains and it showed good results.

B. Classification Rules Construction Using ACO

Ant-Miner proposed by Parpinelli is a rule-based classifier belongs to non-parametric data classification category, which has been successfully applied to different classification problems [8]. However, it can’t cope with continuous attributes directly; so it requires the continuous attributes to be discretized by creating discrete intervals in a pre-processing step. The goal of ACO classifier is to extract classification rules from training data. The algorithm discovered an ordered list of classification rules based on a heuristic function. The classification rules are generally expressed with IF-THEN rule, and each classification rule has the form: IF <term1 AND term2 AND ... -> THEN <class>. Each term is a triple<attribute, operator, value>, where value is a value belonging to the domain of attribute, operator represents a relational operator. A high-level pseudo-code of Ant-miner is shown in Fig. 1.

TrainingSet= [all training cases];
DiscoveredRuleList= []; /* Rule list is initialized with an empty list*/
WHILE (TrainingSet≥Max_Uncovered_Cases)
i=1; /* ant index */
j=1; /* convergence test index */
ELSE j=1;
THEN j= j+1;
IF (R
j
is equal to R
i-1
) /* update convergence test */
ELSE j=1;
END IF
i =i+1;
UNTIL (i=No_of_Ants) OR (j >=No_Rules_Converg)
TrainingSet= TrainingSet- {set of cases correctly covered by R
best
};
END WHILE

Fig. 1. Overview of Ant-miner (Parpinelli et al., 2002)

A cAnt-Miner algorithm was used in our study for land cover classification, which is proposed by Fernando extended from Ant-Miner. The cAnt-Miner has the ability to create discrete intervals for continuous attributes “on-the-fly”, taking advantage of all continuous attributes information, rather than requiring a discretization preprocessing step [9], [10].

The cAnt-Miner approach extended Ant-Miner to support continuous attributes in the rule construction taking the form of (attribute < value) or (attribute ≥ value), where value is a value belonging to the domain of the continuous attribute. It incorporated an entropy-based discretization method into Ant-Miner’s rule construction process to dynamically create thresholds on continuous attributes domain values. Both operator and best value are determined when an ant selects one continuous attribute to construct a classification rule.

1) Heuristic function

Heuristic function is an estimate of the quality of this term, with respect to its ability to improve the predictive precision of the rule. The quality is measured in terms of the entropy for preferring this term to the others, and is given by Equations (1), (2):

\[ \eta_i = \frac{\log_2 (K) - \text{entropy}(T_{ij})}{\sum_j \log_2 (K) - \text{entropy}(T_{ij})} \]  

\[ \text{entropy}(T_{ij}) = -\sum_{k=1}^{K} \left( \frac{\text{freq}_T}{|P|} \right) \times \log_2 \left( \frac{\text{freq}_T}{|P|} \right) \]  

where \( k \) is the number of classes, \( |P| \) is the total number of examples in partition \( T_{ij} \), \( \text{freq}_T \) is the number of cases in partition \( T_{ij} \) with class \( w \), \( a \) is the total number of attributes, and \( b_i \) is the number of values in the domain of attribute \( i \).

The heuristic value associated with each term in cAnt-Miner also involves a measure of entropy, in addition need a threshold value \( V \) to dynamically partition the continuous attribute \( a_i \) into two intervals: \( a_i < v \) and \( a_i ≥ v \). And the best threshold value is the boundary point \( v \) that minimizes the entropy of the partition according to the Equation (3):

\[ e(p_i, v) = \frac{|S_{a_i < v}|}{|S|} \text{entropy}(a_i < v) + \frac{|S_{a_i ≥ v}|}{|S|} \text{entropy}(a_i ≥ v) \]  

where \( |S_{a_i < v}| \) represents the total number of examples in partition \( a_i < v \), \( |S_{a_i ≥ v}| \) represents the total number of examples in partition \( a_i ≥ v \), \( |S| \) represents the total number of training examples. After selection of the best threshold \( v_{best} \), the entropy of \( trem \) corresponds to the minimum entropy value of the two partitions is defined as Equation (4). And only those candidate threshold values form boundaries between classes are evaluated for best value selection. Once the best threshold value \( v_{best} \) and the operator (< or ≥) are selected, a term in the form of triple \((a_i, \text{operator}, v_{best})\) is added to the ant’s current partial rule.

\[ \text{entropy(term)} = \min(\text{entropy}(a_i < v_{best}), \text{entropy}(a_i ≥ v_{best})) \]
An ant adds one term at a time into the rule, according to the value of heuristic function and the amount of pheromone. The probability that \( \text{term}_{ij} \) is chosen to be added to the current rule is given by Equation (5):

\[
P_y = \frac{\tau_y(t) \cdot \eta_i}{\sum_i \sum_j \tau_y(t) \cdot \eta_i}, \quad \forall i \in I
\]

where: \( \eta_i \) is the value of the heuristic function for \( \text{term}_{ij} \), \( \tau_y(t) \) is the amount of pheromone associated with \( \text{term}_{ij} \) at iteration \( t \), \( I \) are the attributes \( i \) not yet used by the ant.

2) Rule Pruning

The rule pruning is to remove irrelevant terms that might have been unduly included in the rule. The rule pruning procedure iteratively removes the term whose removal will cause a maximum increase in the quality of the rule. The quality of a rule is measured using the following equation (6):

\[
Q = \frac{\text{TruePos}}{\text{TruePos} + \text{FalseNeg}} \times \frac{\text{TrueNeg}}{\text{FalsePos} + \text{TrueNeg}}
\]

where: TruePos is the number of cases covered by the rule that have the same class predicted by the rule, FalsePos is the number of cases covered by the rule that have a different class from the class predicted by the rule, FalseNeg is the number of cases that are not covered by the rule, but have the class predicted by the rule, TrueNeg is the number of cases that are not covered by the rule and that do not have the class predicted by the rule.

3) Pheromone Updating

In the cAnt-Miner’s pheromone updating procedure, the pheromone values are associated with continuous attributes not considering the operator and threshold value. And the procedure has two basic ideas, as following:

- The amount of pheromone associated with each \( \text{term}_{ij} \) occurring in the constructed rules is increased, and the pheromone updating for a \( \text{term}_{ij} \) is performed according to equation (7):

\[
\tau_y(t + 1) = \tau_y(t) + \tau_y(t) \cdot Q
\]

- The amount of pheromone associated with each \( \text{term}_{ij} \) that does not occur in the constructed rules is decreased, corresponding to the phenomenon of pheromone evaporation in real ant colony swarms. And in this situation, the pheromone updating for a \( \text{term}_{ij} \) is performed according to equation (8):

\[
\tau_y(t + 1) = \frac{\tau_y(t)}{\sum_i \sum_j \tau_y(t)}
\]

III. EXPERIMENTAL RESULTS

The ACO based algorithm for land cover classification procedure includes remote sensing data pre-processing, feature extraction, and land cover mapping with ACO based algorithm.

A. Remote sensing Datasets and Preprocessing

In this paper, the optical remote sensing dataset and SAR dataset are combined to land cover classification. The study area is mainly located in Beijing city, 116°10’E ~ 116°53’E, 39°38’N ~ 40°11’N, covering an area of about 3636km²(2020x2000pixels). The multispectral remote sensing data was acquired by Landsat5 TM sensor on 6 May 2005 (see Fig.2). The Envisat ASAR images were acquired on April 13, 2005 and December 20, 2005.

Fig. 2. Landsat TM image of Beijing, pseudo colour composite with RGB band 5, 4, 3.

The Landsat TM image is geometrically corrected by selecting reference points based upon a topographic map of the study area with a Universal Transverse Mercator (UTM) map projection. The reference points are evenly distributed in the study area, and the root mean square error (RMSE) of the registration process is less than one pixel.

The Envisat ASAR data selected in this study is level 1B, and the work mode is alternating polarization (AP), and it is precision image. For obtaining backscatter values, two ASAR images were calibrated using BEST (Basic Envisat SAR Toolbox) software. The backscatter images were filtered with Adaptive Frost filter, in a moving 5x5 window. And after the backscatter images were resampled using a nearest-neighbour approach to 30m pixel size, the two images were geo-corrected referring to the Landsat TM image, and the RMSE are less than one pixel.

B. Feature Extraction and Input Datasets

The Principal Component Analysis (PCA) and Tasseled Cap transformation were used as feature extraction methods. We select the first three principal component of Landsat TM as three bands of the input vector. Tasseled Cap Transformation (TCT) is one of the available methods for enhancing spectral information content of Landsat TM data [11]. The Tasseled Cap index was calculated from the related six TM bands are brightness, greenness and wetness. The mean of PC1 with 3x3 window is selected as a texture feature. The total input bands and their descriptions are shown by Table I.

C. Land Cover Classification with ACO Based Algorithm

1) Training dataset and testing dataset

As indicated by the high spatial resolution aerial photos and the field survey, we selected the training dataset and testing dataset with six land cover classes separately. The descriptions of land cover classes are shown in Table II.

The training dataset separability can be measured by the Jeffries-Matusita (JM) distance. In order to evaluate the
quality of training data set, the JM distances are calculated between the pair of land cover classes. The JM distances between the two classes are mostly > 1.9, so we can conclude that the training dataset have good separability, using it to land cover classification is a convincing research work.

### TABLE I: THE INPUT FEATURE VECTOR OF LAND COVER CLASSIFICATION

<table>
<thead>
<tr>
<th>Input bands</th>
<th>Description of land cover feature vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band1</td>
<td>PC1 (the first Principal component of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band2</td>
<td>PC2 (the second Principal component of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band3</td>
<td>PC3 (the third Principal component of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band4</td>
<td>ASAR1 (backscatter index image after filter of Envisat ASAR on April 13, 2005)</td>
</tr>
<tr>
<td>Band5</td>
<td>ASAR2 (backscatter index image after filter of Envisat ASAR on December 20, 2005)</td>
</tr>
<tr>
<td>Band6</td>
<td>TCT1 (brightness, Tasseled Cap Transformation of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band7</td>
<td>TCT2 (greenness, Tasseled Cap Transformation of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band8</td>
<td>TCT3 (wetness, Tasseled Cap Transformation of Landsat TM on May 6, 2005)</td>
</tr>
<tr>
<td>Band9</td>
<td>Mean (the mean texture of PC1)</td>
</tr>
</tbody>
</table>

### TABLE II: DESCRIPTION OF TRAINING DATASET AND TEST DATASET

<table>
<thead>
<tr>
<th>Class number</th>
<th>Type</th>
<th>Description</th>
<th>Training dataset</th>
<th>Test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bare land</td>
<td>farmland covered with soil</td>
<td>804</td>
<td>786</td>
</tr>
<tr>
<td>2</td>
<td>Bare land</td>
<td>airports, squares, other land with high reflectivity</td>
<td>881</td>
<td>360</td>
</tr>
<tr>
<td>3</td>
<td>Urban land</td>
<td>buildings, roads</td>
<td>2041</td>
<td>1150</td>
</tr>
<tr>
<td>4</td>
<td>Farmland</td>
<td>farmland covered with crops</td>
<td>1230</td>
<td>549</td>
</tr>
<tr>
<td>5</td>
<td>Green land</td>
<td>trees, grass</td>
<td>1013</td>
<td>504</td>
</tr>
<tr>
<td>6</td>
<td>Water</td>
<td>reservoirs, lakes, rivers</td>
<td>826</td>
<td>407</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td>6795</td>
<td>3756</td>
</tr>
</tbody>
</table>

2) Classification results

The ACO algorithm was used to construct classification rules based on the selected training dataset, and the parameters of ACO algorithm were set as follows:

- Number of ants (No_of_Ants): 60
- Minimum number of cases per rule (Min_Cases_per_rule): 10
- Maximum number of uncovered cases in the training dataset (Max_Uncovered_Cases): 10
- Number of rules used to test convergence of the ants (No_Rules_Converg): 10
- Number of Iterations: 1500.

The cAnt-Miner algorithm has discovered 12 ordered rules. Since in the ACO algorithm, the ant colony can search the best rule at each iteration time from the training data, so the higher rank rule from the result can get the more accurate land cover classification results than the lower rank ones. And the input feature vectors constructing the high rank rule also show more importance than the lower rank ones. From the rule list, we can see that the rules include Envisat ASAR vectors are almost at the bottom of the list, which shows that the Envisat ASAR data is the ancillary data for the land cover classification, and adding the SAR data always improve the classification accuracy of that land cover class with low separability. Using the rules, the land cover classification map was produced (see Fig. 3).

As can be observed in Table III and Table IV, cAnt-Miner obtained the best overall rank based on the classification precision and Kappa coefficient among the four algorithms being compared. On the other hand, it can be observed that using the combination of multiple features can evidently help the discrimination of bare farmland and urban land. So, using the combined optical and SAR dataset can improve the overall precision.

3) Classification validation and analysis

For comparison purposes, two other well known classifiers in remote sensing data classification research field were used in the classification experiment to evaluate the performance of cAnt-Miner classifier. One is MLC—a probabilistic method, the other is also most well known decision tree method namely C4.5. We also applied the same training dataset to train MLC classifier and C4.5 classifier, and evaluate them using the same test dataset.

Confusion matrices of the results for cAnt-Miner, Ant-miner, MLC and C4.5 using the same test dataset are calculated. Comparing the confusion matrix of cAnt-Miner with MLC, it is seen that using cAnt-Miner classifier can improve the bare farmland and bare land greatly than using MLC classifier, and the precision of other land cover classes changed little. Comparing the confusion matrix of cAnt-Miner with C4.5, it is seen that using cAnt-Miner method can to a large extent improve the classification precision of urban land, farmland and green land than using C4.5 classifier. In general, using ACO classifier has great advantages to land cover classification with combination of the optical and the SAR dataset in comparison of using MLC classifier with C4.5 classifier. The classification from ACO classifier has achieved considerably higher precision than that obtained from MLC classifier and C4.5 classifier. The overall precision obtained from cAnt-Miner classifier is 95.47%, and from Ant-Miner classifier is 95.26%, they are higher than that of MLC and C4.5. And from data mining point of view to comparison, the rules that ACO algorithm has found are much simpler than that found by C4.5. The cAnt-Miner can slightly improve the precision and can construct simpler rules comparing with the Ant-Miner. The number of rules by cAnt-Miner is 12, but the decision tree that constructed by C4.5 are very complex, the tree have 562 leaves, and the size of tree is 593. In summary, using the cAnt-Miner classifier can get high classification precision and also well simplifies the classification rule set.

IV. CONCLUSION AND DISCUSSION

In this paper, the ACO algorithms has been presented for land cover classification, and also the paper demonstrates the usefulness of combined using of the Landsat TM and the Envisat ASAR data covering Beijing area for land cover
classification. The ACO classifier can combine the multiple source spatial datasets to land cover classification, and it can constructed the ordered classification rules that will select the best features extracted from optical and SAR dataset. So the ACO can help improving the discrimination ability among the different land cover classes. So from the experimental results, it is observed that the ACO outperforms the other two algorithms, MLC and C4.5. And the cAnt-Miner classifier can construct much simpler rules and higher accuracy.

In summary, from the experimental observations it is seen that cAnt-Miner classifier is an effective approach for automatic land cover map generation from combining optical and radar remotely sensed images.

<table>
<thead>
<tr>
<th>Class</th>
<th>Bare farmland</th>
<th>Bare land</th>
<th>Urban land</th>
<th>Farmland</th>
<th>Green land</th>
<th>Water</th>
<th>User precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare farmland</td>
<td>739</td>
<td>55</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>92.61</td>
</tr>
<tr>
<td>Bare land</td>
<td>47</td>
<td>295</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>82.63</td>
</tr>
<tr>
<td>Urban land</td>
<td>0</td>
<td>10</td>
<td>1128</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>97.66</td>
</tr>
<tr>
<td>Farmland</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>534</td>
<td>2</td>
<td>0</td>
<td>99.63</td>
</tr>
<tr>
<td>Green land</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>500</td>
<td>0</td>
<td>96.15</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>390</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Produced precision (%) = 94.02, 81.94, 98.09, 97.27, 99.21, 95.82

**Table 3: Confusion Matrix for the Classification Results of ACO Using Combination of TM and ASAR Dataset**

**Table 4: Comparison of Classification Obtained From ACO, MLC and C4.5**

<table>
<thead>
<tr>
<th>classifier</th>
<th>Overall precision (%)</th>
<th>Kappa coefficient</th>
<th>Percent of misclassification (%)</th>
<th>Simplicity of rule sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAnt-Miner</td>
<td>95.47</td>
<td>0.9435</td>
<td>4.53</td>
<td>Number of rules: 12</td>
</tr>
<tr>
<td>Ant-Miner</td>
<td>95.26</td>
<td>0.9410</td>
<td>4.74</td>
<td>Number of rules: 31</td>
</tr>
<tr>
<td>C4.5</td>
<td>92.92</td>
<td>0.9119</td>
<td>7.08</td>
<td>Number of leaves: 562</td>
</tr>
<tr>
<td>MLC</td>
<td>90.31</td>
<td>0.8797</td>
<td>9.69</td>
<td>Size of tree: 593</td>
</tr>
</tbody>
</table>

**Acknowledgment**

The authors thank for the cAnt-Miner software shared by Fernando E. Otero.

**References**


Qin Dai received the Ph.D. degree in the institute of remote sensing application (IRSA), Chinese academy of sciences (CAS). She is currently an associate professor of Institute of Remote Sensing and Digital Earth (RADI), CAS. Her research interests include machine learning, artificial intelligent methodologies applied to remote sensing image classification, image retrieval, and image change detection. She is a referee for several international journals.

Shibing Liu is an associate professor of Institute of Remote Sensing and Digital Earth (RADI), CAS. His research interests include spatial database and spatial data services.

Jin Yang is an associate professor of Institute of Remote Sensing and Digital Earth (RADI), CAS. His research interests include remote sensing image retrieval and spatial data services.

Zhaoming Zhang received his Ph.D. degree in Cartography and Geographical Information System in institute of remote sensing applications, CAS (University of Chinese Academy of Sciences). His current research interests are land surface parameters retrieval and remote sensing applications.