Research Article

An Associate Rules Mining Algorithm Based on Artificial Immune Network for SAR Image Segmentation

Mengling Zhao^{1,2} and Hongwei Liu¹

¹*School of Mathematics and Statistics, Xidian University, China*

²*School of Science, Xi'an University of Science and Technology, China*

Correspondence should be addressed to Mengling Zhao; zhaomengling@126.com

Received 15 January 2015; Revised 26 May 2015; Accepted 28 May 2015

Academic Editor: Hak-Keung Lam

Copyright © 2015 M. Zhao and H. Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As a computational intelligence method, artificial immune network (AIN) algorithm has been widely applied to pattern recognition and data classification. In the existing artificial immune network algorithms, the calculating affinity for classifying is based on calculating a certain distance, which may lead to some unsatisfactory results in dealing with data with nominal attributes. To overcome the shortcoming, the association rules are introduced into AIN algorithm, and we propose a new classification algorithm an associate rules mining algorithm based on artificial immune network (ARM-AIN). The new method uses the association rules to represent immune cells and mine the best association rules rather than searching optimal clustering centers. The proposed algorithm has been extensively compared with artificial immune network classification (AINC) algorithm, artificial immune network classification algorithm based on self-adaptive PSO (SPSO-AINC), and PSO-AINC over several large-scale data sets, target recognition of remote sensing image, and segmentation of three different SAR images. The result of experiment indicates the superiority of ARM-AIN in classification accuracy and running time.

1. Introduction

In data mining field, classification is one of the most important issues which have a wide range of applications. The aim of classification can be defined to classify an object into a predetermined group or a category according to some features observed in itself. In recent years, many classification methods have been widely applied to pattern recognition, machine learning, data mining, and artificial intelligence. So far, a large number of classification algorithms have been proposed, including the K -nearest neighbor (KNN) algorithm, Bayesian classification, Fuzzy C -means (FCM), decision tree method, and BP neural network algorithm. However, these traditional methods have their drawbacks in dealing with classification problems. Therefore, new methods have been proposed in recent years, and one of the most successful methods is artificial immune classification system inspired by biological immune system.

Artificial immune network has a large number of applications since it was developed. In 1996, the Cooke net [1] was proposed by Hunt and Cooke. They used the net in

DNA sequence for pattern recognition. The Multiple-Valued Immune Network (MVIN) [2] proposed by Tang et al. not only simulates the B cells but also makes use of the auxiliary and inhibition of T cells. The main idea in this model is the simulation of the interreaction and regulation between B cell and T cell with a learning mechanism of a multivalued feature set. Then the evolutionary immune network was first employed in data analysis [3] by de Castro and von Zuben. As a simulation of the antibody cloning and selection process in biological immune system, the method regards the data set as the antigen set and generates the output as the antibody set. By the analysis of the simulation results, we find that although the algorithm is hard to determine the number of clusters, even in the case of knowing the number of clusters which may deviate from the correct cluster center, the algorithm can greatly reduce the data redundancy, and the resultant antibody set can reflect the structural features of the data set. In 2001, they proposed a new model, aiNet [4]. The model differs from Hunt-coke mostly in the absence of consideration of the concept of stimulation and the model only considers the affinity. To obtain the data pattern, some

stages of the training process are driven by clonal selection, and some remove redundancy through the interaction of antibody. In 2012, a segmentation algorithm which is based on the negative selection mechanism of the artificial immune system was proposed in [5]. Experiments show that the algorithm has a good performance on the segmentation threshold subjectively. Recently, Fu et al. proposed a coordinated immune network template algorithm [6], and it added a coordination mechanism between innate immunity and adaptive immunity in order to extract targets from blurred infrared images. More recently, an image segmentation method for blurred trace infrared images was proposed in [7], which was considered the function of immune factors for blurred infrared image segmentation. The simulation results show that this method can improve the target segmentation rate and reduce the segmentation error rate. Resource Limited Artificial Immune System (RLAIS) developed by Timmis and Neal uses Artificial Recognition Ball (ARB) to represent classified cells. All of the ARBs compete with each other based on the incentive level to which they are subjected. The higher the incentive level of ARB is, the more B cells it can represent, and one ARB may be pruned away from the network for its incapacity of standing for any of the B cells. For its good control over the population size, Watkins and Timmis made a series of improvements of this algorithm [8–10], and de Castro and von Zuben proposed Self-Stabilizing Artificial Immune System (SSAIS) [11] on the basis of RLAIS for continuous analysis of time-varying data. Nevertheless, SSAIS differs from RLAIS quietly in the case of no supply of the resource constraints level, and each ARB in it can control resource on its own level (decentralized control). Moreover, SSAIS computes the stimulation level without the consideration of the inhibition among B cells. Later, an improving version of SSAIS called Metastable Memory Immune Network (MSMIN) emerged. The algorithm, which clings to the idea of AINE, is applied to data analysis, clustering, and artificial immune memory. More recently, artificial immune network has turned out to be a hot research area, which is proved by the stably growing number of related conferences and journal articles [12–19].

Existing artificial immune network classification algorithms are used to find the affinity of immune cells through the calculation of the distance, which determines the classification result far from ideal when the characteristics of the data to be classified are nominal characteristic value. Since the algorithm is a supervised search for the best cluster center, it is more difficult to deal with high-class classification problems. The convergence time of the algorithm is not linearly related to the number of classification categories; with the increase of number of classification categories, the time of calculation increases, while the accuracy of classification decreases. When the number of classification categories grows to a certain extent, the algorithm will result in an incapable classification (the classification result is poor or the calculation time is too long). In order to make the algorithm more versatile and robust, we made some improvements on the basis of the new artificial immune network algorithm [12] and introduced the association rules. Association rules have developed a good application in data mining algorithms.

Do et al. [20] implemented a favorable combination between associated classification and artificial immune system clonal selection algorithm, which lead to a new classification algorithm called artificial immune system-associative classification (AIS-AC), and the method had advantage over the traditional association rules classification methods in running time and classification performance. Inspired by this, we introduced the associate rules into the artificial immune network and improved the associate rules mining algorithm based on artificial immune network (AIN-ARM) for large-scale data, remote sensing image, and SAR image classification.

2. Background

2.1. Idiotypic Immune Network Theory. Following Bernet's clonal selection theory, in 1974, Jerne proposed the idiotypic and anti-idiotypic of immune system internal regulation of the idiotypic immune network theory. For his research work in this area, he won the Nobel Prize in 1984. Idiotypic immune network theory suggests that, before the stimulation of the body from antigens, it is still in a relatively stable statement; on the entrance of antigens, it destroys the balance and results in the generation of the specific antibody molecules. Once the number of the molecules reaches a fixed value, the idiotype immune response of the anti-immunoglobulin molecules takes place; namely, the anti-idiotypic antibody are produced. Therefore, the antibody molecules can be identified by their anti-idiotypic antibody molecules just as they recognize antigens. This view is accepted whether by the antibody molecules in the blood or the immunoglobulin (Ig) molecules which act as the antigen receptors on the surface of lymphocytes. In the same animal's antibody molecules, a set of idiotype determinants on the antibody molecules of one animal's body can be identified by another group of anti-idiotypic antibody molecules, which is the same as a set of antigen receptor molecules on the surface of the lymphocytes that can be recognized by another group of anti-idiotypic antibody molecules on the surface of lymphocytes. In this way, a network structure consisting of lymphocytes and antibody molecules is formed in the body. The network theory states that the production of this type of anti-idiotypic antibody plays an important role in the regulation of immune response for the clones proliferated by antigen stimulation to be inhibited rather than endlessly proliferated and for a stable balance of the immune response to be maintained.

2.2. Associate Rule. Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of data items, D is a set of transaction data (called transaction database), each transaction T of D is a subset of data items of I , that is, $T \subseteq I$; each transaction T has a unique identification number called TID. Let A be a set of data items, if and only if $A \subseteq T$, being called transaction T contains A . An association rule is shaped like an " $A \Rightarrow B$ " implication, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \emptyset$. The associate rule $A \Rightarrow B$ can be described in the following parameters:

- (1) Support: $\text{sup}(A \rightarrow B) = P(AB)$

- (2) Confidence: $\text{conf}(A \rightarrow B) = P(B | A) = \frac{\text{sup}(AB)}{\text{sup}(A)}$.

Here the sup is the percentage of transactions in D which contains both A and B . As a measurement of the importance of the association rules, sup explains how much the representative value of this association is. The higher the support is, the more important the association rules are. If item set meet the minimum support, then call it the frequent item set. Conf refers to the percentage of transactions in D consisting of A as well as B , and, as the measurement of accuracy of association rules, it reflects the posterior probability of the occurrence of B subject to the premise of a given A .

3. The Proposed Algorithm

Based on the biological immune network model and the associated classification, we proposed a new classification algorithm, an associate rules mining algorithm based on artificial immune network (AIN-ARM); the algorithm used the association rules to represent the memory cells of artificial immune network, and it also used mining association rules as a replacement of the search of the existing artificial immune network algorithm for the best cluster center. The algorithm has strong robustness and versatility and a convergence time which is linearly proportional to the classification categories, so the problem that the number of the classification categories is too high to classify is nonexistent. The immune academic language that the algorithm refers to and its meaning in this algorithm are presented as follows:

- (i) Immune cell: the associate rule.
- (ii) Memory cell: the association rules which are selected for the final classification.
- (iii) Antigen Ag : Ag is defined as the training samples which is carried in the artificial immune network algorithm, denoted by $Ag = (Ag_1, Ag_2, \dots, Ag_C)$, and Ag_i is the i th collection of training samples.
- (iv) Clone size: P_c represents the size of immune cell proliferation. The number of clone is proportional to the affinity value, if the affinity value of the i th immunocyte is denoted by $\text{affinity}(i)$, and the number of clone is $\text{affinity}(i) * P_c$.
- (v) Mutation rate: in this paper, P_m is the mutation probability of each cloning individual.
- (vi) Affinity value: affinity value indicates the matching degree of the immune cell to the antigen and is represented as affinity. In this algorithm, the affinity is the confidence value.
- (vii) Minimum support: it is a threshold condition to distinguish whether a rule is a frequent rule or not and denoted by minsupport . Let the support of the i th rule be $\text{support}(i)$, if and only if $\text{support}(i) \geq \text{minsupport}$, and the rule is useful.
- (viii) Minimum confidence: it is denoted by minconfidence . If and only if a frequent rule meets the minconfidence , that is, $\text{confidence}(i) \geq \text{minconfidence}$ (where the

$\text{confidence}(i)$ denotes the rule's confidence value), the rule is called association rule.

This section first presents the key technologies in the algorithm in detail, including data preprocessing, network initialization, antibody affinity function, hypermutation, pruning operations, suppression of the immune network operation, and classified operation, and the steps of the algorithm are given in the end.

3.1. Key Technologies of Algorithm

3.1.1. Data Preprocessing. Of the input data, the continuous features should be discretized. Data can be preprocessed according to

$$E' = \left\lceil \frac{E - \min(E)}{\max(E) - \min(E)} \cdot L \right\rceil. \quad (1)$$

Here, E means the input data to be processed; $\min(E)$ and $\max(E)$ represent, respectively, the minimum and maximum of a column to which a continuous feature data belongs. L is the range size of data discretization, which represents the right rounding.

3.1.2. Network Initialization. Immune network initialization includes the generation of the candidate rule set, the initialized population of immune cells, and the weight vector of W . The rules candidate set contains all the association rules which have a length of 2 and is similar to " $X_i \Rightarrow C_j$ " (here, X_i represents a characteristic value of a certain dimension; C_j is the category property value). Immune cells initialization includes that some association rules are selected randomly as the initial immune cells from the candidate rule set to the network. Set the initial confidence value of each immune cell for the value of its initial weight vector, and weight vector has the following functions: (1) When the network is suppressed, it promotes the retention of high-quality memory cells. (2) When it is classified, it is used to evaluate the value of membership grade between the test samples and the memory cells.

3.1.3. Function of Antibody Affinity. The affinity value of this algorithm is the value of confidence. We assume that an immune cell B_i such as $X_1 \wedge X_2 \wedge \dots \wedge X_n \Rightarrow C_i$, then the affinity value is calculated as

$$\begin{aligned} \text{affinity}(B_i) &= \text{conf}(B_i) \\ &= P(C_i | X_1 X_2, \dots, X_n) \\ &= \frac{\text{sup}(X_1 X_2, \dots, X_n C_i)}{\text{sup}(X_1 X_2, \dots, X_n)} \end{aligned} \quad (2)$$

$$\text{sup}(X_1 X_2, \dots, X_n) = P(X_1 X_2, \dots, X_n),$$

where $\text{sup}(X)$ represents the obtaining of the support value of X , which is the occurrence probability of event X .

3.1.4. Super Variation. Since association rules are unique, the variation mentioned by this algorithm only acts on a certain

position of immune cells (change the value of the position or add a new value to the position); the new value is chosen from the candidate rules set randomly, so it guarantees that the new value meets the minimum support value and minimum confidence value. Once the offspring has different values on two or more positions with the parent, the two rules would generally be irrelevant, and then the offspring would lose the inheritance of the parent. The mutation probability (p_m) is positively related to the immune cell affinity.

3.1.5. Pruning Operation. In the process of evolutionary search, it will produce a lot of redundant rules. To improve the efficiency of this algorithm, these rules should be deleted. The cancel rules include: duplicate rule sets, long rules with a low confidence value, and rules contained in the memory cells of immune networks.

3.1.6. Suppression Operation of the Immune Network. To improve the efficiency of the algorithm, it is necessary to eliminate parts of the memory cells with the elite selection

principle. The elimination is obtained by mutual inhibition among memory cells. The network suppression rule is that the memory cell A is eliminated when it is contained in the memory cell B ($A \subseteq B$), when and only when the weight vector corresponded with A is not greater than that with B ($W_A \leq W_B$).

3.1.7. Classification Operation. Associated classification algorithms generally use maximum degree of confidence as the basis for classification. It means that, while in the classification process, a test sample should travel through all the association rules and choose all the rules which are capable of covering the test sample, and then the sample belongs to the category which has the rule with the largest confidence value. This classification approach combines the training and classification process of association rules well together. However, direct application of this algorithm may produce the problem that some samples fail to be classified (because only some relatively good association rules can be chosen as the memory cells). Therefore, classification rules of the algorithm are set as

$$d = \begin{cases} w & z = x \\ w \cdot 10^{\wedge}(-y) \cdot \left(10^{\wedge} \left(2 \cdot \frac{(y-1)}{(x-1)} \right) \cdot \frac{(L - \text{distance})}{L} \right)^{\wedge} z & 0 < z < x \\ 0 & z = 0, \end{cases} \quad (3)$$

where d is the degree of membership between M and T , M represents memory cells, x is its length, W is its weight value (that is, confidence value), T means the test sample, y is the largest length of memory cell, L indicates the dispersion, z is the number of the same features between M and T , and distance expresses the difference between the two features.

The flowchart of the proposed algorithm is shown in Algorithm 1.

4. Experimental Results and Analysis

In order to verify the effectiveness of the proposed ARM-AIN, the performance of the ARM-AIN is compared with other four state-of-the-art methods. They are artificial immune network classification algorithm (AINC), artificial immune network classification algorithm based on Self-adaptive Particle Swarm Optimization (SPSO-AINC) and artificial immune network classification algorithm based on Particle Swarm Optimization (PSO-AINC) and K-NN (KNN) algorithm.

The experiments include three parts. In the first part, we test the performance of all compared algorithms with four data sets with large size. In the second part, five algorithms are applied to a target recognition problem of remote sensing image. In the third section, we apply five algorithms to segment three synthetic aperture radar (SAR) images. In all experiments, we have taken 50 independent runs (with different seeds of the random number generator) of each algorithm.

The encoding mechanism in AIN-ARM is modified on the basis of an artificial immune network classification (AINC) algorithm, and each immune cell includes all the category information in the antigen; therefore the proposed algorithm is more suitable for dealing with data sets with large number of categories. In SPSO-AINC, an adaptive weight is introduced to the AINC mutation operator, which is specified as follows:

$$V(i, j) = \text{rand} * p_m * |MC(i, j) - \text{train}(k, j)| * (MC(i, j) - \text{train}(k, j)) \quad (4)$$

$$\begin{aligned} MC'(i, j) &= MC(i, j) + W(i) * V(i, j) + r_1 * \text{rand} \\ &* \exp(-f(i)) + r_2 * \text{rand} \\ &* (gBest(j) - MC(i, j)). \end{aligned} \quad (5)$$

In which, w is the adaptive weight, MC represents the immune cells before variation, MC' means the immune cells after variation, $\text{train}(k, j)$ expresses the j -dimension elements of training sampler randomly selected with the same category in the mutation fragments, MC indicates the group cloned, f is the fitness value, the global extreme $gBest$ represents the individual with the highest affinity degree, $MC(i, j)$ means the j -dimensional elements of the i th individual selected, and r_1 and r_2 represent learning parameters of which the sum is 1.

In PSO-AINC, a basic PSO mutation is adopted; other one is similar to that of AINC.

Step 1. Read and preprocess (discrete the continuous feature) the data.

Step 2. Set the minimum support and confidence ($minSupport$, $minConfidence$), maximum rule length max_long , clone size p_c , mutation probability p_m and so forth; meanwhile initialize immune network net_M , weight vector W , set the initial value of i to 1, $i \leq C$ (C is the number of categories of data).

Step 3. Network evolution

3.1: Generate candidate rule set $candidate_i$ of the i th training samples, (which contains all the rules that have a length of 2, meets the minimum support value rule), and choose a certain amount of rules with the highest confidence value from candidates as the initial immune cells to be added into M_i , set the number of iterations: iteration = 0, coverage: coverage = 0, the number of generation with stable coverage: no_change = 0;

3.2: Calculate the affinity values of immune cells in M_i , and sort them in descending order by affinity values;

3.3: Clone MC_i from the immune cells of M_i , clone number and the affinity is proportional;

3.4: Super variate each individual in MC_i ; each individual only variates on one position, the group variated is noted MCM_i ;

3.5: Calculate support and confidence of each individual in MCM_i ; and add the individuals which meet the minimum support and confidence rule and cover at least one or more (includes one) training samples in the category as the memory cells to net M ; then remove these immune cells from MCM_i , add their corresponding confidence values to the weight vector W , and update coverage;

3.6: Replace the original immune cells in M_i with the immune cells in MCM_i , then M_i performs the pruning operation;

3.7: net_M performs the network suppression;

if iteration is an integer multiple of some integer,
then
 reduce the minimum confidence value, select parts of rules randomly from $candidate_i$ for adding to M_i in order to maintain the diversity of immune cells;
end if

if the value of coverage increases,
then
 no_change = 0;
 iteration = iteration + 1;
else
 iteration = iteration + 1;
 no_change = no_change + 1;
end if

if (coverage == 1 || iteration == 50 || no_change == 10)
then
 i = i + 1;
 if i > C
 then
 go to Step 4;
 else
 go to Step 3.1;
 end if
end if

Step 4. Output network

ALGORITHM 1: Association rule mining algorithm based on artificial immune network.

Main parameters employed in these algorithms are set as follows.

AIN-ARM. Mutation probability $p_m = 0.8$, clone size $p_c = 10$, the maximum number of iterations is 50, the range of the minimum support $minsupport$ is (0.001, 0.1), and the initial minimum confidence $minconfidence$ is 1.

SPAINC. Mutation probability $p_m = 0.8$, clone size $p_c = 10$, the maximum number of iterations is 100, and the stop condition is that the affinity of the best individual is 1 or as follows.

AINC. The parameter setting is the same with SPAINC.

The stop condition is set that the total coverage of training samples of the category reaches 100%, or it is not changed within 10 generations.

All the algorithms discussed here have been developed in a Matlab7.0.1 platform on a Pentium-IV 2.33-GHz PC, with a 2-GB main memory in Windows Server 2003 environment.

4.1. Experimental Results on Large-Scale Standard Data Sets.

The properties of the test data sets including Letter, Nursery, Adult, and Digit [21] are shown in Table 1. The last two columns of Table 1 represent the number of training samples and test samples selected from different data sets, respectively.

Table 2 gives experimental results including average classification accuracy and average running time (CPU) over

TABLE 1: Basic property of data sets used in experimental section.

Data	Data size	Continuous features	Discrete features	Term characteristics	Number of categories	Number of data missing some attributes	Training samples	Test samples
Letter	20000	0	16	0	26	0	13333	6667
Nursery	12960	0	0	8	5	0	8640	4320
Adult	48842	6	0	8	2	3620	32561	16281
Digit	10992	16	0	0	10	0	7494	3498

TABLE 2: Experimental results of five algorithms on large-scale data sets.

Classification algorithm		Letter	Nursery	Adult	Digit
SPSO-AINC	Classification accuracy	46.8%	58.2%	81.9%	83.4%
	Running time	$1.42e + 05$	$9.48e + 03$	$1.73e + 04$	$1.28e + 04$
ARM-AIN	Classification accuracy	62.3%	93.2%	83.2%	88.2%
	Running time	$5.32e + 04$	$2.17e + 03$	$1.50e + 04$	$4.74e + 03$
AINC	Classification accuracy	46.9%	57.5%	80.7%	78.2%
	Running time	$8.12e + 05$	$2.13e + 04$	$2.41e + 04$	$2.89e + 05$
PSO-AINC	Classification accuracy	47.1%	60.7%	79.7%	78.9%
	Running time	$1.22e + 05$	$1.04e + 04$	$1.78e + 05$	$2.84e + 05$

50 independent runs, and the running time in Table 2 is in seconds (s). All experimental results is obtained by a 10-fold crossmatching technology.

As shown in Table 2, for Letter, it has a large category number (26), so there is a great difference in classification results. AIN-ARM increases by 15% compared to SPSO-AINC, and it requires the least running time compared to other methods. Since the category number of Nursery is 5 and all of its features are nominal, ARM-AIN increases by approximately four times in running times compared with SPSO-AINS and obtains 35.3% higher classification accuracy than SPSO-AINS and it is best among five algorithms. We can see ARM-AIN has a slight improvement in Adult in terms of classification accuracy compared with other four algorithms. For Adult, we can see the proposed algorithm is still best whether in classification accuracy or running time; however, the improvement is not very obvious; the reason is that Adult has two categories. For Digit with a relatively larger classification category number, ARM-AIN gets an obvious improvement in classification accuracy and running time compared with other algorithms.

4.2. Target Recognition of Remote Sensing Image. In this section, we apply five methods to solve target recognition of remote sensing image. Experimental images are real measured remote sensing images. Each contains only the target and background. The entire images set are 1064 pieces of images which have a variety of targets with different rotating angles, scales, and incomplete parts. Their basic condition are aircraft class 608 and ship class 456, and parts of images are shown in Figure 1; aircraft training takes 160 samples as training data set, we extract 120 from ship data sets as training samples, and the rest ones are used for testing.

FIGURE 1: Parts of remote sensing images.

Since it is widely used that the seven constant features of two-dimensional images have the features of rotation, translation, and scale invariance, the algorithm still extracts seven invariant moments as the classification features [20].

Two-dimensional image $f(x, y)$ has a $(p + q)$ central moment:

$$u_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y), \quad (6)$$

where $\bar{x}$ and $\bar{y}$ represent the average value of the horizontal and the vertical coordinates of all pixels in the images and $f(x, y)$ expresses the gray value of the pixel located in (x, y) .

In this case, seven constant moments M_1 – M_7 based on second- and third-order central moments are shown as

$$\begin{aligned}
M_1 &= (u_{20} + u_{02}) \\
M_2 &= (u_{20} - u_{02})^2 + 4u_{11}^2 \\
M_3 &= (u_{30} - 3u_{12})^2 + (3u_{21} - u_{03})^2 \\
M_4 &= (u_{30} + u_{12})^2 + (u_{21} + u_{03})^2 \\
M_5 &= (u_{30} + u_{12})(u_{30} - 3u_{12}) \\
&\quad \cdot [(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2] + (3u_{21} - u_{03}) \\
&\quad \cdot (u_{21} + u_{03}) [3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2] \\
M_6 &= (u_{20} - u_{02}) [(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2] \\
&\quad + 4u_{11}(u_{30} + u_{12})(u_{21} + u_{03}) \\
M_7 &= (3u_{21} - u_{03})(u_{30} + u_{12}) \\
&\quad \cdot [(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2] - (u_{30} - 3u_{12}) \\
&\quad \cdot (u_{21} + u_{03}) [3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2].
\end{aligned} \tag{7}$$

To obtain the scale invariance, equation needs to be handled as follows, where P is the number of pixels contained in the target $r = (u_{20} + u_{02})$:

$$\begin{aligned}
M'_1 &= \frac{M_1}{P}; \\
M'_2 &= \frac{M_2}{r^2}; \\
M'_3 &= \frac{M_3}{r^3}; \\
M'_4 &= \frac{M_4}{r^2}; \\
M'_5 &= \frac{M_5}{r^6}; \\
M'_6 &= \frac{M_6}{r^4}; \\
M'_7 &= \frac{M_7}{r^6};
\end{aligned} \tag{8}$$

classification results of remote sensing image are shown in Table 3. We can see that ARM-AIN is better in all three measures than SPSO-AINC. The former increases by 2.1% compared with the latter in classification accuracy. Its standard deviation also decreases by 2.1, which means ARM-AIN performs more stably. The two algorithms are far different in training time: AIN-ARM takes no more than 11 s, yet SPAINC takes about 1760 s. The proposed algorithm is far better than other three algorithms in terms of average classification accuracy and running times over 50 independent runs.

TABLE 3: Classification results of remote sensing image.

Classification algorithm	Average classification accuracy	Standard deviation	Running time (in seconds)
SPSO-AINC	91.5%	7.5	1.76e + 03
ARM-AIN	93.6%	5.4	10.61
AINC	90.8%	13.6	2.48e + 03
PSO-AINC	89.8%	15.4	1.85e + 03

4.3. Experimental Results on SAR Image Classification. To reflect the effectiveness of ARM-AIN further, this section selects three SAR images to evaluate the performance of the proposed algorithm.

For the good characteristics performed in texture features analysis of the statistic amount based on GLCM and the wavelet transformation, this study extracts 4-dimensional characteristics [22] based on GLCM, respectively, which are called angle second-order moment (energy), contrast, entropy and correlation, and seven-dimensional features in which there are two-layer wavelet extract energy of offspring. Therefore, what the algorithms deal with is 11-dimensional texture feature value of two-layers wavelets transforming and GLCM.

The first image is an X-band SAR subimage of a Swiss lake, its size being 140×155 , as shown in Figure 2(a), which contains three types of ground objects: lakes, city, and mountains. Various types of representative areas are selected as training samples, which are 200, 100, and 200, respectively. We get the test samples of lake, city, and mountains, respectively, as 225, 121, and 300. Both of the sliding window sizes of GLCM and wavelet are 7. Figures 2(b), 2(c), and 2(d) are classification results of four algorithms, respectively.

As seen from Figure 2, all of the algorithms can basically achieve correct classification of the lake and the city, and the overall classification result of ARM-AIN is better compared to other algorithms. AINC, SPSO-AINC, and PSO-AINC misclassify more pixel points which belong to the mountain. Moreover, ARM-AIN has more obvious superiority in running time which takes only 5.98 s while AINC and SPAINC take 41.65 s and 22.13 s, respectively.

Table 4 presents confusion matrix obtained by the five algorithms, and Table 5 gives average classification accuracy and kappa coefficient of the five algorithms, which indicates that AIN-ARM is best among five algorithms and, especially, it increases by 0.3% and 1.6%, respectively, compared with SPSO-AINC and AINC in classification accuracy and increase by 0.005 and 0.0239, respectively, in the Kappa coefficient.

The size of the second SAR image is 256×256 , which contains four ground objects: river and three kinds of crops. It can be seen from Figure 3(a) that the number selected as training samples of water, white, gray, and light crop is 100, 100, 200, and 200, respectively. GLCM and wavelet sliding window sizes are set as 5 and 13, respectively. Figures 3(b)–3(d) are separately classification results of four algorithms.

Figure 3 shows that ARM-AIN remains better than other three algorithms in the edge areas with four crosscutting

FIGURE 2: Classification results of five algorithms on the first SAR image.

TABLE 4: Comparison of confusion matrix of four methods.

Algorithm		Lake	City	Mountain	Total
AIN-ARM	Lake	225	0	0	225
	City	0	114	4	118
	Mountain	0	7	296	303
	Total	225	121	300	635
SPSO-AINC	Lake	225	0	0	225
	City	0	120	12	132
	Mountain	0	1	288	289
	Total	225	121	300	633
AINC	Lake	225	0	0	225
	City	0	120	20	140
	Mountain	0	1	280	281
	Total	225	121	300	625
PSO-AINC	Lake	225	0	0	225
	City	0	120	22	142
	Mountain	0	1	278	279
	Total	225	121	300	623

TABLE 5: Comparison in the average classification accuracy and the Kappa coefficient of four algorithms.

Accuracy	ARM-AIN	SPSO-AINC	AINC	PSO-AINC
Average classification accuracy	98.52%	97.01%	95.89%	96.21%
Kappa coefficient	0.9715	0.9591	0.9319	0.9411

ground objects; other three algorithms mistake mutually more seriously and both of them have more misclassifications between gray and light crops; the former mistakes lots of pixels that belong to gray crops, while the latter is just the opposite. However, ARM-AIN is a bit weaker in regional consistency of light crops classification because of some noises. ARM-AIN, AINC, and SPAINC take running time of 28.32 s, 288.13 s, and 258.68 s, respectively. It shows that ARM-AIN is of great superiority in convergence rate more than nine times as much as the later ones.

FIGURE 3: Classification results of four algorithms on the second SAR image.

As shown in Tables 6 and 7, ARM-AIN increases by 4.4%, 4.2%, and 7.6% in average classification accuracy compared with SPSO-AINC, AINC, and PSO-AINC, respectively. Whether from the view of classification accuracy or kappa coefficient, the proposed algorithm is superior to the other three algorithms.

The size of the third SAR image is 256×256 , which includes four kinds of surface features: water, urban, and two crops. Referring to Figure 4(a), 225, 64, 200, and 200 samples are selected separately from each type of surface features as the training samples, and the test samples are 200, 64, 200, and 200, respectively. The sizes of gray level cooccurrence

FIGURE 4: Classification results of four algorithms on the third SAR image.

TABLE 6: Comparison of confusion matrix of four methods.

Algorithm		Water	Gray	Light	White	Total
AIN-ARM	Water	90	0	0	0	90
	Gray	10	197	14	0	221
	Light	0	3	186	31	220
	White	0	0	0	69	69
	Total	100	200	200	100	542
SPSO-AINC	Water	88	0	0	0	88
	Gray	12	199	28	0	239
	Light	0	1	172	34	207
	White	0	0	0	66	66
	Total	100	200	200	100	525
AINC	Water	90	0	0	0	90
	Gray	10	147	1	0	158
	Light	0	53	199	23	275
	White	0	0	0	77	77
	Total	100	200	200	100	513
PSO-AINC	Water	99	0	0	0	99
	Gray	1	199	0	0	200
	Light	0	1	105	0	106
	White	0	0	95	100	195
	Total	100	200	200	100	503

TABLE 7: Comparisons in the average classification accuracy and kappa coefficient of four algorithms.

Item	ARM-AIN	SPSO-AINC	AINC	PSO-AINC
Average accuracy	90.27%	86.85%	86.02%	82.59%
Kappa coefficient	0.8712	0.8297	0.7986	0.7475

matrix and sliding window are 7. The classification results are shown in Figures 4(b), 4(c), and 4(d) respectively.

It can be seen from the classification results in Figure 4 that ARM-AIN has obvious advantage over SPSO-AINC, PSO-AINC, and AINC in the results of water, and the latter misclassified more pixels of the water to the dark crops. For both crops, the classification results of ARM-AIN are most ideal, SPSO-AINC misclassified more pixels of light crops to the dark crops, and AINC is opposite. Overall, the classification results of ARM-AIN are best. For running time, the consumption time of AIN-ARM, AINC, and SPAINC is 15.25 s, 237.23 s, and 188.92 s, respectively, and the first one has a time less than one-tenth of the latter two. It can be seen clearly that ARM-AIN plays a notable role in convergence rate.

FIGURE 5: Gaussian noise with different parameters and the corresponding classification results.

As can be seen from Tables 8 and 9, ARM-AIN obtained higher average classification accuracy than SPSO-AINC, PSO-AINC, and AINC, respectively.

4.4. Noise Sensitivity of the Proposed Method. In order to discuss the noise sensitivity of the proposed algorithm, we take the first SAR image as an example and add two kinds of noise, that is, Gaussian noise and salt noise, into the original SAR image then adopt the proposed algorithm to segment the image with noise. The noise parameters are set as 0.02, 0.05, and 0.1, respectively, and Figures 5 and 6 show the results of classification.

As shown in Figure 7, with the parameter value increasing, the classification accuracy decreases; compared to the Gaussian noise, Salt noise makes the classification results worse, due to Gaussian noise having an evenly distribution in the image, which is also proved by the classification accuracy provided in Table 10.

4.5. Analysis of Algorithm Parameters. There are five main parameters used in the proposed algorithm, that is,

TABLE 8: Comparison of confusion matrix of four methods.

Algorithm		Light	Water	Gray	Urban	Total
AIN-ARM	Light	200	0	9	5	214
	Water	0	165	0	0	165
	Gray	0	35	191	0	226
	Urban	0	0	0	59	59
	Total	200	200	200	64	615
SPAINC	Light	200	0	61	0	261
	Water	0	144	0	0	144
	Gray	0	56	139	0	195
	Urban	0	0	0	64	64
	Total	200	200	200	64	547
AINC	Light	197	0	98	0	295
	Water	0	116	5	0	121
	Gray	3	84	97	0	184
	Urban	0	0	0	64	64
	Total	200	200	200	64	474
PSO-AINC	Light	195	0	34	0	229
	Water	0	165	49	0	214
	Gray	5	35	117	0	157
	Urban	0	0	0	64	64
	Total	200	200	200	64	541

FIGURE 6: Salt noise with different parameters and the corresponding classification results.

FIGURE 7: The relation between the classification accuracy and *minsupport*.

TABLE 9: Comparison in the average classification accuracy and kappa coefficient of four algorithms.

Item	ARM-AIN	SPSO-AINC	AINC	PSO-AINC
Average accuracy	92.86%	81.94%	72.08%	81.46%
Kappa coefficient	0.9013	0.7531	0.6004	0.7424

TABLE 10: Comparison in the average classification accuracy and kappa coefficient of four algorithms.

Item	0.02	0.05	0.1
Gaussian noise	95.52%	89.89%	83.76%
Salt noise	94.17%	87.26%	81.33%

the initial size of immune cells *num*, the clone size p_c , the minimum support *minsupport*, and the minimum confidence *minconfidence*. Where the initial size of immune cells *num*, clone size p_c are connected directly with the time complexity of the algorithm and when the other parameters remain

unchanged, the greater the size of the antigen population and proportion of clone are, the better the population diversity is, which is at the expense of the algorithm's time complexity. Minimum confidence of the confidence of the algorithm is adaptive; it will gradually reduce with the number of iterations increasing, so the minimum confidence level has little effect on the algorithm, and the initial value can be set to 1. For the mutation probability p_m , in the general immune algorithm, it always has a great impact on the performance of the algorithm; however, the unique mutation operator and pruning operations in the ARM-AIN algorithm cause p_m a little impact on the performance of the algorithm; while it only concerns the convergence time of the algorithm, the larger mutation probability can lead to a faster convergence rate, and in this paper we set $p_m = 0.8$.

The main consideration in the algorithm design is the effect of parameter *minsupport* on the performance. The size of the minimum support is related to the number of the frequent rules to be searched; the smaller the value of the *minsupport* is, the more the frequent the rules are; the greater the searching space is, the right higher the classification accuracy of the algorithm is; however, when the *minsupport* reduces to some extent and confronts the problem of frequent rule explosion, the convergence time becomes unacceptable too. To sum up, the reasonable range of *minsupport* is from 0.01 to 0.1. Here we use the data of Digit as the example and sample the parameter *minsupport* in the range from 0.01 to 0.11 with the step size 0.01 and examine the effect of the classification accuracy. The results are shown in Figure 7.

5. Conclusion

Based on the biological immune network theory and associated classification, we proposed an association rules mining algorithm based on artificial immune network (ARM-AIN), and the proposed algorithm is applied to large-scale data mining, target recognition of remote sensing image, and classification of SAR images. The algorithm can solve multiclass or nominal-attribute classification problem which cannot be solved well by most of existed immune network classification algorithms. Moreover, AIN-ARM is superior in the results of SAR image classification to SPSO-AINC and AINC and particularly dominates in the convergence rate. Consequently, all the above illustrate the algorithm plays a good role in its versatility and robustness.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] J. E. Hunt and D. E. Cooke, "Learning using an artificial immune system," *Journal of Network and Computer Applications*, vol. 19, no. 2, pp. 189–212, 1996.
- [2] Z. Tang, T. Yamaguchi, K. Tashima, O. Ishizuka, and K. Tanno, "Multiple-valued immune network model and its simulations," in *Proceedings of the 27th International Symposium on Multiple-Valued Logic*, pp. 233–238, Antigonish, Canada, May 1997.
- [3] L. N. de Castro and F. J. von Zuben, "An evolutionary immune network for data clustering," in *Proceedings of the 6th Brazilian Symposium on Neural Networks (SBRN '00)*, pp. 84–89, IEEE, Rio de Janeiro, Brazil, 2000.
- [4] L. N. de Castro and F. J. von Zuben, "Immune neural network models: theoretical and empirical comparisons," *International Journal of Computational Intelligence and Applications*, vol. 1, no. 3, pp. 239–257, 2001.
- [5] D. Fu, X. Yu, and T. Wang, "Segmentation algorithm study for infrared images with occluded target based on artificial immune system," in *Proceedings of the 8th International Conference on Computational Intelligence and Security (CIS '12)*, pp. 350–353, Guangzhou, China, November 2012.
- [6] D. Fu, X. Yu, and H. Tong, "Target extraction of blurred infrared image with an immune network template algorithm," *Journal of Optics and Laser Technology*, vol. 56, pp. 102–106, 2014.
- [7] Y. Yu, "Blurred trace infrared image segmentation based on template approach and immune factor," *Journal of Infrared Physics and Technology*, vol. 67, pp. 116–120, 2014.
- [8] A. Watkins, J. Timmis, and L. Boggess, "Artificial immune recognition system (AIRS): an immune-inspired supervised learning algorithm," *Genetic Programming and Evolvable Machines*, vol. 5, no. 3, pp. 291–317, 2004.
- [9] A. Watkins and J. Timmis, "Artificial immune recognition system (AIRS): revisions and refinements," in *Proceedings of the ICARIS*, pp. 173–181, Springer, 2002.
- [10] A. Watkins and J. Timmis, "Exploiting parallelism inherent in AIRS, an artificial immune classifier," in *Artificial Immune Systems: Third International Conference, ICARIS 2004, Catania, Sicily, Italy, September 13–16, 2004. Proceedings*, vol. 3239 of *Lecture Notes in Computer Science*, pp. 427–438, Springer, Berlin, Germany, 2004.
- [11] L. N. de Castro and F. J. von Zuben, "Artificial immune systems: a novel approach to pattern recognition," in *Artificial Neural Networks in Pattern Recognition*, pp. 67–84, University of Paisley, Renfrewshire, UK, 2002.
- [12] R. Liu, M. Niu, and L. Jiao, "A new artificial immune network classifier for SAR image," in *MIPPR 2009: Pattern Recognition and Computer Vision*, vol. 7496 of *Proceedings of SPIE*, pp. 1–8, Yichang, China, October 2009.
- [13] W. Li, H. Huang, C. Wang, and H. Tang, "Synthetic fault diagnosis method of power transformer based on rough set theory and improved artificial immune network classification algorithm," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, pp. 676–681, IEEE, Jinan, China, October 2008.
- [14] C. Danzhen, A. Qian, and C. Chen, "The application of artificial immune network in load classification," in *Proceedings of the 3rd International Conference on Deregulation and Restructuring and Power Technologies (DRPT '08)*, pp. 1394–1398, April 2008.
- [15] A. Bhaduri, "University time table scheduling using genetic artificial immune network," in *Proceedings of the International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom '09)*, pp. 289–292, Kottayam, India, October 2009.
- [16] M. Khaleghi, M. M. Farsangi, H. Nezamabadi-pour, and K. Y. Lee, "Voltage stability improvement by multi-objective placement of SVC using modified artificial immune network algorithm," in *Proceedings of the IEEE Power & Energy Society General Meeting (PES '09)*, pp. 1–7, July 2009.

- [17] Y. Zhong, L. Zhang, B. Huang, and P. Li, "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 420–431, 2006.
- [18] F. O. de Franca, F. J. Von Zuben, and L. N. de Castro, "An artificial immune network for multimodal function optimization on dynamic environments," in *Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO '05)*, pp. 289–296, New York, NY, USA, 2005.
- [19] A.-L. Chen and Q. Guo, "An effective hybrid optimization algorithm based on self-adaptive particle swarm optimization algorithm and artificial immune clone algorithm," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, pp. 129–132, IEEE, October 2008.
- [20] T. D. Do, S. C. Hui, A. C. M. Fong, and B. Fong, "Associative classification with artificial immune system," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 217–228, 2009.
- [21] C. L. Blake and C. J. Merz, UCI Repository of Machine Learning Databases, <http://archive.ics.uci.edu/ml/datasets.html>.
- [22] D. A. Clausi and B. Yu, "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 215–228, 2004.