

Analysis of Classify Remote Sensing Images Based on Artificial Bee Colony Algorithm

Jing Chen, Hongmei Huang, Xiaoxuan Wang, and Yujing Wu

Huali College Gangdong University of Technology, Gangzhou Zengcheng, China

Keywords: artificial bee colony; bio-inspired computing; remote sensing; classification, feature selection, multiple classifier system.

Abstract: This paper proposed the feature selection and optimization of support vector machine parameters based on artificial bee colony algorithm. The goal of algorithm is to improve the classification rate and reduce the number of features. The proposed algorithm is validated on the UCI database. The result show that the proposed algorithm has the ability of feature selection and it can improve the classification rate obviously. Compared with other algorithms, the proposed algorithm is more efficient in feature selection and classification.

1. Introduction

Remote sensing image, especially high-resolution remote sensing image, has rich spatial information, such as geometry, structure, texture etc. Traditional pixel-based classification method, which utilizes only spectral information, not only leads to the wastage of spatial information, but also affects the effectiveness of classification seriously, because of the phenomenon of same objects having different spectral character and different objects having same spectral character and the existing of noises [1]. In recent years, more and more people research on high-dimension data in data mining. Classification is an effective method for data analysis. However, the ever-increasing dimension of the data including some irrelevant or redundant features increases the difficulty of classification. So, the selection of feature subsets and optimization of parameters of classifier are two important factors to improve the classification performance for the classification problem [2].

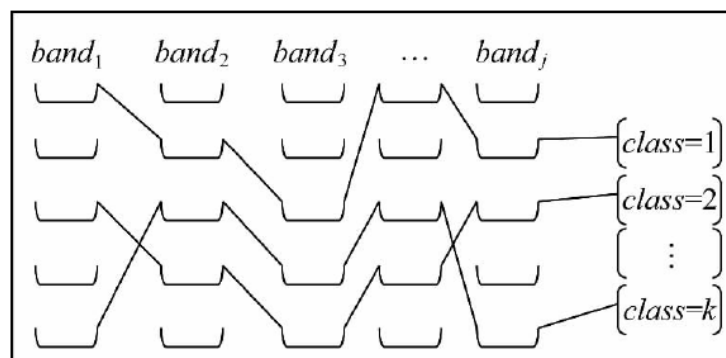


Fig. 1 Remote sensing classification rule based on ABC

As a new algorithm, intelligent optimization algorithm can be used to solve the complex practical problems by simulating the group behavior. For instance, artificial bee colony algorithm can be used to solve the problems by simulating the intelligent foraging behavior of honeybee. The algorithm not only makes the local searching efficient but also has the ability of global optimization with the help of heuristic search strategy. Based on the advantages of simple implementation and strong robustness, it has been used in pattern recognition and intelligent control.

2. Feature Selection and Parameter Synchronization

When using ABC algorithm to optimize the problem, we should first determine what parameters to be optimized and how to design it to achieve the desired effect. In this paper, we mainly use ABC algorithm to select feature subset and optimize the parameters of SVM. When optimizing the parameters for SVM, we should first determine which parameters need to be optimized and the selection of different kernel functions needs to optimize the different parameters. Therefore, the type of the kernel function should be determined before the parameters are optimized[3].

For RBF kernel functions, the parameters include two: penalty factor C and kernel function parameter γ . Take different values, you will get different classification results. The penalty factor C is used to control the complexity of the model and the penalty of misclassification of the sample [4, 5]. When C is too large, there will be a phenomenon of high classification accuracy in the training phase and poor test phase; if C is too small, the obtained result is difficult to satisfy. The parameter γ will affect the distribution of the sample in the high-dimensional space. It determines the complexity of the classification surface, and the influence of the parameter γ on the classification accuracy of the SVM is much greater than the effect of the penalty factor C on the classification. The problem of "learning" is too small and there will be "less study". Therefore, optimizing C and γ plays an important role in improving classifier performance.

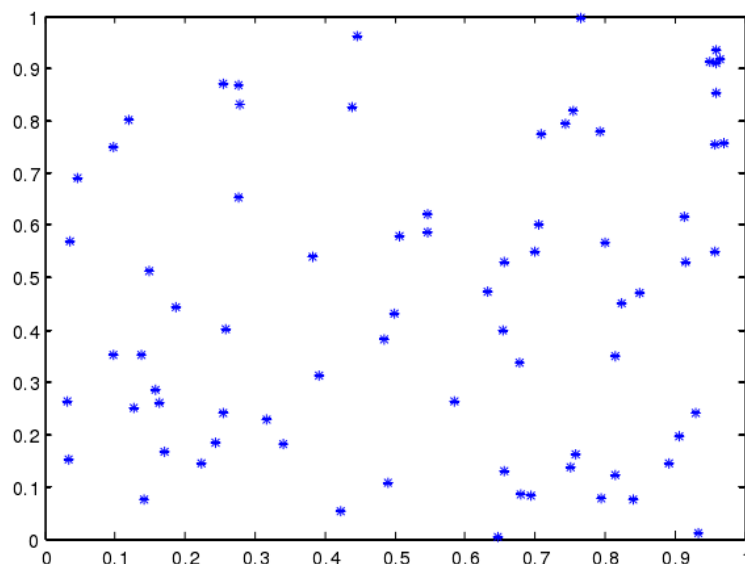


Fig. 2 Solution generation of standard ABC algorithm

About the data of experiments, in order to enable the end up with the experimental results, the ideal before the experiment, all need to deal with data, using different methods of data processing, the result will be different. In particular, there is singular data (too large or too small data will appear in the data). In order to eliminate the differences between data, the original data must be processed. The original data after pretreatment, can bridge the gap between data, in classifying data or other operation experiment, can shorten the operation time and at the same time to speed up the convergence of algorithm. There are many ways to normalize data. In practice, it is necessary to consider whether or not to apply data standardization, and which kind of data standardization should be applied, depending on the specific circumstances.

3. Feature Selection Based on Artificial Bee Colony Algorithm

In order to make the results more effective, the K -fold cross validation was set to 5 in the experiment. The data are randomly divided into five disjoint subsets, each of which is basically the same size. One subset is a test set and the other four subsets are used as a training model. The program runs five times. Each subset is used as a test data set, and the accuracy rate of each running test is

calculated. The total number is divided by five, and the classification accuracy of the data set in this experiment is obtained. Cross validation is characterized by independent test sets and reliable results.

$$fitness = w_a \times svm_accuracy + w_f \times N^{-1} \quad (1)$$

The proposed algorithm runs on the data set in the UCI database, which shows the feasibility of the algorithm. In order to reflect the superiority of the algorithm, it is necessary to compare the algorithm with other algorithms to evaluate the algorithm.

$$f: x \rightarrow y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

In the standard ABC algorithm, both the initial solution and the solution to the bee phase are randomly generated by formula (2).

The solution sets obtained are too scattered and random, which will affect the quality of the optimal solution and the convergence speed of the algorithm. In practical applications, an optimization problem will appear multiple targets, and these goals are contradictory, that is to say, improving the performance of some targets will weaken the performance of other targets and can not achieve the best of all targets at the same time. Therefore, the multi-objective optimization problem is to get the best solution between the numerous objectives. The corresponding mathematical model is shown as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

The fitness function curve on the data set is shown in Figure 3, and the real line is an improved algorithm. The dotted line is an unimproved algorithm. The improved algorithm reaches the optimal value at a faster convergence speed, improves the quality of the optimal solution and speeds up the running speed of the algorithm.

Table 1 UCI data

| No. | data | score | distribution |
|-----|------------|-------|--------------------|
| 1 | Iris | 3 | 30.40.50 |
| 2 | Glass | 6 | 70.76.17.13.3.9.29 |
| 3 | Ionosphere | 1 | 225.162 |
| 4 | Wdbc | 2 | 531.221 |
| 5 | wdbc | 4 | 511.74 |

Table 2 Iris

| Iris | Setosa | Versicolou r | Virginic a |
|-----------------|--------|-----------------|---------------|
| Setosa | 15 | 0 | 0 |
| Versicolou r | 0 | 16 | 0 |
| Virginica | 0 | 0 | 17 |

From Table 2, we can see that when the ABC algorithm is used to optimize feature subsets and SVM parameters, the results obtained are significantly better than the results of the standard ABC algorithm. The improved algorithm is better than the one obtained in the previous chapter. For example, the data sets with significantly improved classification rate in the two-category dataset have WDBC and WPBC, and the classification rates have increased by 7.39% and 7.90%, respectively; while the dataset with significantly improved classification rate in multi-classification problems is glass, increasing from the original 69.44% to 75.00%. Although the degree of improvement in the classification rate of the remaining data sets is not very obvious, the results have also improved.

From the classification results, it can be seen that when the multi-objective algorithm is used to

improve the performance of the classifier, multiple sets of optimal solutions will appear. The obtained results, some of the obtained classification rates will be better than the results obtained by the single target, and some classification rates. Although it will be slightly inferior to the single-target classification rate, the number of features obtained will be less than that obtained by a single target, and the decision maker may select a result with a high classification rate or a small number of features as required. According to the classification results in the table, the results of the training set and the test set are not much different, indicating that the multi-objective algorithm has better classifier parameters optimized, and the number of features is also significantly reduced, especially for the data sets Ionosphere, WDBC and WPBC with a large number of features.

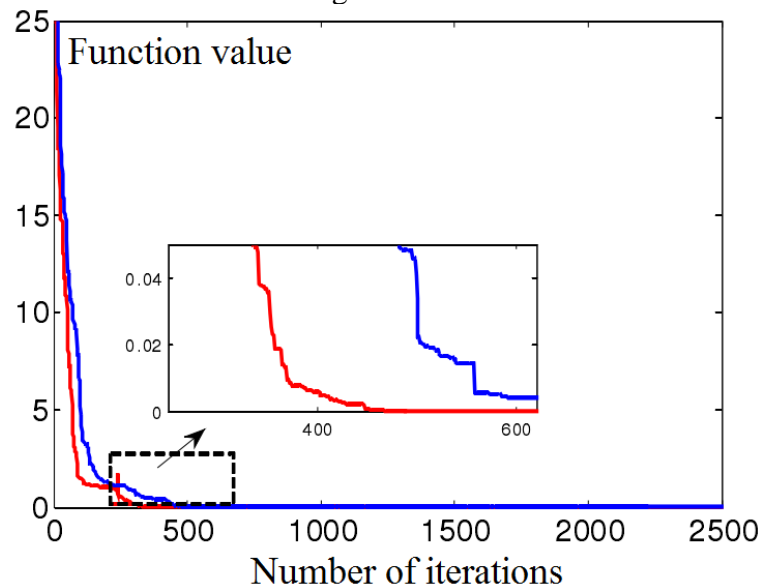


Fig. 3 Function convergence curve

4. Summary

By the combination of the bio-inspired computing methods, an intelligent method to classify remote-sensing images using artificial bee colony algorithm is proposed. This ABC-based classification method searches for the optimal split points on each band of remote-sensing image automatically, where the optimal solution of split points emerge from intelligent behaviour of honey bee swarms. The method defines bees search path as the connection between optimal split points and feature category node, and constructs classification rules in the form. This smart approach can express the complex classification more clearly without mathematical formulas. By taking an example of ALOS image in the north shore of the Yangtze River estuary, the proposed classification method based on ABC algorithm were implemented.

References

- [1] Jayanth J, Koliwad S, Kumar T A: Classification of remote sensed data using Artificial Bee Colony algorithm (Egyptian Journal of Remote Sensing & Space Sciences, 2015), 18(1), p.119-126.
- [2] Min C, Shi Z, Yang J: An Innovative Method to Classify Remote Sensing Images Using Artificial Bee Colony Algorithm (Acta Geodaetica Et Cartographica Sinica, 2013).
- [3] Sun X, Yang L, Zhang B, et al: Hyperspectral image clustering method based on Artificial Bee Colony algorithm (Sixth International Conference on Advanced Computational Intelligence. IEEE, 2014) p.095047.
- [4] Chandrakala D, Sumathi S: Image Classification based on Color and Texture features using FRBFN network with Artificial Bee Colony Optimization Algorithm (International Journal of

Computer Applications, 2014), 98(14) p.19-29.

[5] Sun X, Yang L, Zhang B, et al: An Endmember Extraction Method Based on Artificial Bee Colony Algorithms for Hyperspectral Remote Sensing Images (Remote Sensing, 2015), 7(12) p.16363-16383.