Analysis of Remote Sensing Image Classification Method Based on Genetic Algorithm

Jing Chen

Huali College Gangdong University of Technology, Gangzhou Zengcheng, China

Keywords: genetic algorithms, object-oriented, feature selection, multiple classifier system.

Abstract: Based on the above characteristic, firstly genetic algorithms, which have been widely used in optimization problems, was applied to feature selection techniques, and then a multiple classifier system model was designed as extension of genetic-algorithm-based feature selection techniques. As there are correlations among features, a new multiple classifier system model was constructed with disjoint feature subspace constraint. These two models were trained by supervised methods and oriented by improving the classification accuracy. The construction of multiple classifier system is completed by genetic algorithms, which determines what optimized feature subspace each sub-classifier should uses.

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]. Object-oriented classification techniques of remote sensing image, integrating information of spectrum, shape and texture of image, utilizes spatial information to a great extent so that to meet the drawback of the traditional method while dealing with classification of remote sensing image.
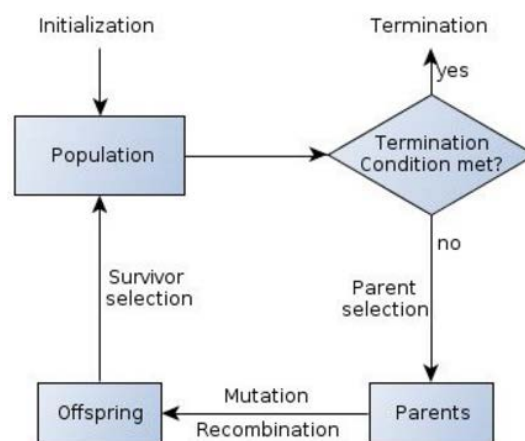

Fig. 1 A flow chart of object-based remote sensing classification

Remote sensing data have been applied in a wide range of fields, such as environmental monitoring, major disaster management, urban planning and national defense strategy [2]. With the development of remote sensing technology and the improvement of satellite spatial resolution, the application of high resolution remote sensing images is more and more extensive. In most remote sensing applications, the ultimate goal is to analyze and interpret images. Classification problem is one of the most fundamental problems in remote sensing image information analysis.

2. Feature Selection and Multiple Classifier System

In a pattern recognition system, feature extraction and feature selection is usually between object features two links between data collection and classification and recognition, the extraction and selection, strongly affect the performance of the classifier. If the pattern features are well selected, and the extracted and selected features show great difference to different types of patterns, we can easily design a classifier with high performance. Therefore, feature extraction and selection are two key technologies in pattern recognition [3].

To evaluate feature subset, we need to measure the effectiveness of feature classification based on a specific criterion [4]. The possible combinations of several characteristics from the high dimensional original feature concentration are many, and which combination is most favorable to the classification and requires a comparative standard. According to the different models of feature selection, the evaluator is also different. Usually, the filter model adopts the criterion of category separability, while the encapsulation model takes the classification accuracy of the encapsulated classification algorithm as the basis of evaluation.
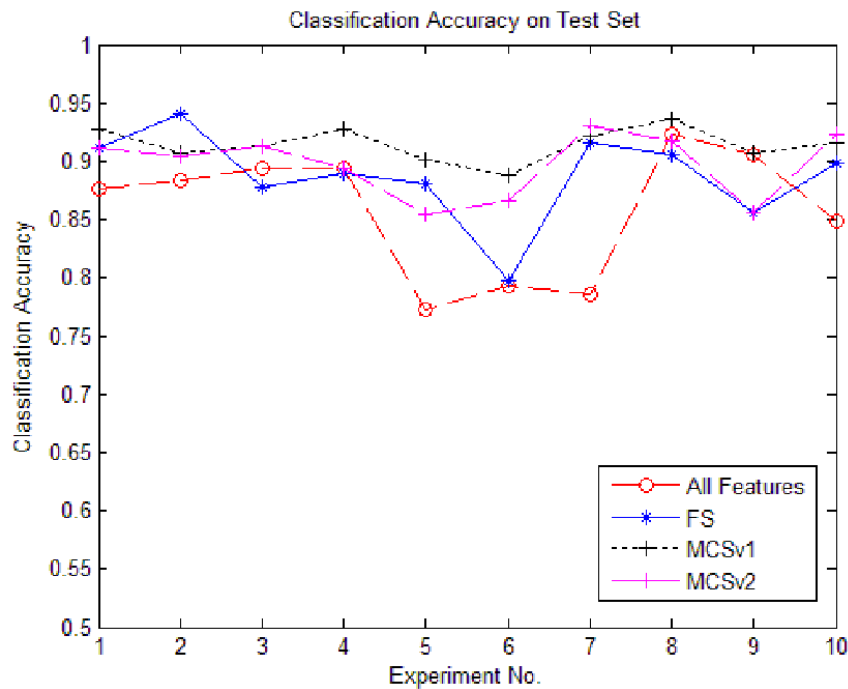

Fig. 2 Contrast of classification accuracy on Test Set of WDBC Dataset

In general, the patterns of different classes can be distinguished because the class domain in the feature space they belong to is a different area. Obviously, the smaller the overlap of these areas is, the better the separability of categories is. Therefore, the class separability criterion can be constructed using geometric distance or disparity measure. This method is mainly how to express the distance between classes. The distance criterion does not consider the probability distribution of various types, but directly calculates the distance from various types of samples. We hope that the intra-class dispersion is as small as possible, and the dispersion between classes is as large as possible.

3. Feature Selection and Multiple Classifier System Based on Genetic Algorithm

In the package model, in order to estimate the classification accuracy, it is known that the sample must be divided into two parts of the training sample and the test sample according to a certain rule. The two part should correspond to the feature subset produced by the feature subset generator. The classifier is trained by the training sample, then the test sample is classified, and the classification results are obtained. A certain accuracy evaluation is defined, that is, the classification accuracy criterion of the feature subset.

$$q = C_D^d = \frac{D!}{(D-d)!d!} \tag{1}$$

The simplest bottom-up search method selects one feature from features that have never been selected so that it has the largest criterion value when combined with features that have been selected. Since the method considers the correlation between the selected features and the features that have been selected, it is generally better than the above single optimal feature selection method. The main disadvantage of the method is that once a feature is selected, even one of the unselected features One is better than it, and it cannot be removed again.

$$D : R^n \to [0, 1]^c \tag{2}$$

Genetic algorithm maintains P(T), which is composed of several individuals (T represents genetic algebra), and each individual represents a potential solution to the problem. Each individual is evaluated by the evaluation function. Some individuals have to undergo random transformations called genetic operations to generate new individuals. There are two main types of transformation: mutation is the method to change the individual to get new individuals; the method of hybridization is to combine the related parts of the two individuals to form a new individual. The evaluation function continues to be evaluated by the evaluation function. From the parent population and the offspring population, the better individuals were selected to form new populations. After several generations of evolution, the algorithm converges to an optimal individual, which is most likely to represent the most generation or sub optimal solution of the problem.

$$f(x) = \begin{cases} C_{\max} - g(x) & g(x) < C_{\max} \\ 0 & g(x) \geqslant C_{\max} \end{cases} \tag{3}$$

Binary coding is the most commonly used encoding method in genetic algorithm. It USES binary symbol set {0, 1} to encode, which has the following advantages: simple coding and decoding operation; Crossover, mutation and other genetic operators are easy to implement. In many combinatorial optimization problems, both the objective function and the constraint function are discrete functions. Usually, the encoding method of feature selection problem is binary coding, because feature selection problem can be regarded as a kind of 0-1 integer rule problem.

From table 1 can see, using GA - BSSVM band selection, the number of features from the original 200 greatly reduced to 101, the classification accuracy is increased from 91.28% to 92.89%, can be seen, on the one hand, the GA - BSSVM method has better performance of band selection; GA - BSSVM, on the other hand, the characteristics of the optimal selection of at the same time, according to the classification accuracy adjustment of parameters in order to determine the optimum parameters of the SVM kernel: C and Y, respectively 157. 89 and 0. 243, it has to do with experiments in this paper, the SVM classification RBF kernel parameter USES the default set of parameter values 100 and 1 get the classification accuracy of 91. 28% are obviously different, and the accuracy is significantly higher than ENIV, LIBSVM and cross validation method for SVM parameters corresponding to the SVM classification accuracy, thus it can be seen, This article adopts the genetic algorithm optimization at the same time band and SVM parameter optimization selection method (GA - BS SVM) is effective, less for band selection and parameter optimization of the classification accuracy has improved significantly, band selection for SVM parameters choosing and image are achieved better optimization results, and the synchronization is complete, the entire implementation process is done automatically, without human to set the parameters according to the results of the experience and training. In terms of parameter selection, the effect of this method is better than that of cross-validation method, and the classification accuracy is better than the default setting of svM classification method in LIBSVM and ENVI.

Table 1 The composition of chromosomes

F=200	C=20	Y=20
F1...fk...fnk	C1...ci...cnc	Y1...yj...ynj

The analysis of Figure 3 shows that the accuracy of image classification increases with the increase of evolutionary algebra, with a larger increase between the twentieth and the fortieth generations, and a more stable growth between the fortieth and the 100th generations, and maintains a relatively high stationary state after the increase of the classification accuracy of the 180th generation to 92.7%. It can be seen that with the continuous evolution, the adaptive function can be gradually close to the ideal value, but the largest evolutionary algebra in this study is 300, the maximum can only evolve to the highest precision of 92.89% after the 260th generation, and the accuracy of the image classification is high, which can meet the requirements.

Fig. 3 Remote sensing image of some area in Mongolia from Spot 5

In summary, the genetic algorithm achieves a good effect of simultaneously optimizing the bands of the hyperspectral image and the parameters of the SVM classifier. The experimental results show that the GA achieves automatic band selection for classification of hyperspectral remote sensing images, compared with the classification accuracy before optimization. The classification accuracy has been improved, which indicates that the GA-B SS VM is an effective method for optimizing the auto-optimized bands and parameters of hyperspectral remote sensing images.

4. Summary

Object-oriented classification needs to use a variety of high-dimensional features with significant differences in nature. Therefore, a single classifier can not achieve a stably ideal result while dealing with those features. Classifier fusion techniques of multiple classifier system have been a hot topic in recent years. In response to the characteristic mentioned above, classifier fusion based on feature selection techniques should be a possible solution.

Experiments show that two models, especially the latter, could effectively improve classification accuracy of object-oriented remote sensing image with high-dimensional features.

References

[1] Guo Y, Wu Y, Ju Z, et al: Remote sensing image classification by the Chaos Genetic Algorithm in monitoring land use changes (Mathematical & Computer Modelling, 2015), 51(11–12), p.1408-1416.

[2] Guo Y N, Xiao D W, Yang M: Genetic Algorithm-Based Support Vector Classification Method for Multi-spectral Remote Sensing Image (Final Program and Book of Abstracts of the 2010 International Conference on Life System Modeling and Simulation & 2010 International Conference on Intelligent Computing for Sustainable Energy and Environment. 2010), p.213-220.

[3] Guo Y N, Xiao D W, Yang M: Genetic Algorithm-Based Support Vector Classification Method for Multi-spectral Remote Sensing Image (Life System Modeling and Intelligent Computing. Springer Berlin Heidelberg, 2017).

[4] Guo Y N, Xiao D W, Yang M: Genetic Algorithm-Based Support Vector Classification Method for Multi-spectral Remote Sensing Image (Life System Modeling & Intelligent Computing, 2016), 6328, p.213-220.