Retraction

Retraction: Research on Classification of Remote Sensing Images Based on Artificial Intelligence (J. Phys.: Conf. Ser. 2074 012034)

Published 9 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 9 September 2022
Research on Classification of Remote Sensing Images Based on Artificial Intelligence

Yuanyuan Peng,*, Jie Liu

1School of Information Engineering, China University of Geoscience Beijing, Beijing, China, 100083
*Corresponding author e-mail: pengyuanyuan@cugb.edu.cn

Abstract. With the rapid development of image processing technology, remote sensing technology has received increasing attention. Relying on artificial intelligence technology and using the advantages of principal component analysis (PCA) to reduce the dimensionality of features, this paper proposes a remote sensing image classification method based on SVM. First, LBP operator is used to extract remote sensing image features, and then PCA is used to perform remote sensing image features. The dimensionality reduction process reduces the feature dimensionality and eliminates feature redundant information, and obtains features that have a large contribution to the classification result. Finally, SVM is used for remote sensing image classification. The results show that PCA-SVM improves the efficiency and accuracy of remote sensing image classification.

Keywords: Support Vector Machine, Classification, Multi-Source Data, Model Selection

1. Introduction

With the rapid development of image processing technology, remote sensing technology has received increasing attention. Through remote sensing technology, remote sensing data can be obtained quickly, efficiently, accurately, and dynamically, and various targets can be classified and identified. The application has become a hot topic in current research [1-4]. However, the classification and recognition of remote sensing images is easily affected by subjective consciousness, and the classification features are not obvious enough. Because some ground features have very similar spectral characteristics, it may be difficult to identify using a single source of remote sensing data [5-9]. Since the distribution of surface features is closely related to the topography, and the topography is determined by the topography, geographic assistance data can help identify these features [10].

This paper proposes a remote sensing image classification method (PCA-SVM) based on artificial intelligence technology. First, LBP operator is used to extract remote sensing image features, and then PCA is used to reduce the dimensionality of remote sensing image features, reduce the feature dimension and eliminate feature redundant information, and obtain features that contribute greatly to the classification results. Finally, SVM is used for remote sensing image classification.
2. Principles of automatic classification of remote sensing images

The automatic classification of remote sensing images is a pattern classification problem, and divides the image into different categories of areas such as lakes, residential areas, rivers, etc. according to different attributes.

Suppose a remote sensing image contains k categories, which are represented by $\omega_i, i=1, 2, \ldots, k$. In order to classify the remote sensing image, k discriminant functions $g_1(x), g_2(x), \ldots, g_k(x)$ must be found, so that each pixel can be classified according to the following rules:

$$x \in \omega_i, \text{if } g_i(x) > g_j(x), i=1, \ldots, k, i \neq j$$

(1)

The principle is shown in Figure 1. Preprocessing is to denoise, cloud and geometrically correct remote sensing images; feature extraction to obtain feature vectors reflecting various targets; training is to learn remote sensing image features to generate training templates for various features; use training templates to perform image processing Classification, according to the principle of automatic classification of remote sensing images. This paper uses PCA and support vector machines for feature extraction and classifier design.

Figure 1. The working principle of automatic classification of remote sensing images

3. SVM

3.1. Two types of SVM classifiers

The working mechanism of the two types of SVM can be summarized as: by looking for a classification hyperplane to make the two types of sample points in the training sample can be separated, and as far as possible from the plane, $H$ is the optimal classification surface, $H_1$ and $H_2$ The distance $m$ between the two categories is called the classification margin.

Where $x \in \mathbb{R}^n$ is the input vector, and $y_1 \in \{1, -1\}$ are two types of problems The SVM method is to find the optimal classification surface $w \cdot \Phi(x) + b = 0$ between the two classes, $b \in \mathbb{R}$ is the bias, and $\Phi$ means the sample The set $S$ is mapped to a high-dimensional feature space.

The process of SVM finding the optimal classification surface is equivalent to solving a convex quadratic optimization problem, Equation (2):

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \xi_i$$

s.t. $y_i [w \cdot \Phi(x_i) + b] \geq 1 - \xi_i$

$$\xi_i \geq 0 \quad i = 1, 2, \ldots, l$$

(2)

Using Lagrangian optimization method and duality principle to solve the problem, Eq. (1), we get Eq. (2):
\[
\min \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{l} \alpha_i \\
\text{s.t.} \sum_{i=1}^{l} y_i \alpha_i = 0 \\
0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, l
\]

Among them, \( K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \) is the kernel function, and the following optimal classification surface with the largest interval can be obtained:

\[
\sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j) + b = 0
\]

The final decision function is:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j) + b \right)
\]

### 3.2. Kernel function

(1) Polynomial kernel function:

\[
K(x_i, x_j) = (\gamma x_i \cdot x_j + b)^d, \quad d = 1, 2, \ldots
\]

(2) Radial basis function (RBF) kernel function:

\[
K(x_i, x_j) = \exp \left( -\gamma \| x_i - x_j \|^2 \right)
\]

(3) Sigmoid kernel function:

\[
K(x_i, x_j) = \tanh \left( \gamma (x_i \cdot x_j) + b \right)
\]

Among them, \( \gamma, \ b, \ d \) are the parameters of the kernel function, and the choice can only be made based on experience.

### 3.3. Multi-class SVM

SVM itself is used to solve two types of problems. There are currently two types of multi-type methods: one is to integrate many two types of problems, and there are already more mature algorithms, such as one-against-one (OAO), one For multi-methods (one-against-all, OAA) and directed-acyclic-graph (DAG) support vector machines, etc.; the other is to directly consider multiple types of problems in the optimization formula, and its disadvantage is the algorithm. The selected objective function is too complicated to be realized, and the amount of calculation is large, so this article uses the former to experiment.

The one-to-many method is the earliest and most widely used multi-type strategy. Each two-class classifier is used to distinguish. The final decision result is that the class with the largest decision value wins. Since it is one class versus the rest, a significant imbalance in the number of samples is likely to occur, leading to certain errors.

The one-to-one method includes \( N(N-1)/2 \) two-class SVM classifiers. Each trained classifier is used to distinguish two categories, and all \( N(N-1)/2 \) two-class classifiers use voting to make decisions. A total of \( N(N-1)/2 \) two-class SVM classifiers are trained. The difference is that in the testing phase, the DAG method uses a bipartite directed acyclic graph starting from the root. A binary decision is
adopted, that is, according to the output value of the left node or the right node, until the leaf node, and finally the leaf node determines the final attribution.

4. Classification
Multi-source information fusion and classification must go through the following steps: (1) Preprocessing of data from various sources, including the conversion of geographic auxiliary data from vector to raster, and various corrections to remote sensing images; (2) Accurate registration of data from different sources; (3) Use a suitable classifier for classification; (4) Perform evaluation and further processing.

4.1. Test area and data
The test area is located in a county, and the remote sensing data uses Landsat ETM images. The size of the test area is 512×512 pixels. The TM data of the 6 bands are: TM-1 to TM-5 and TM-7. The image was generated on March 7, 2020. The additional geographic assistance data includes three data channels: digital elevation model (DEM), aspect and slope.

Because vegetation chlorophyll has strong absorption characteristics at a wavelength of 0.69 µm, the ratio or linear combination of infrared and near-infrared reflectance can realize the expression of vegetation information, and the normalized difference vegetation index (NDVI) is widely used. Applied to the quantitative study of vegetation cover, the normalized vegetation index makes the ground surface with vegetation cover and the ground surface without vegetation cover more distinguishable. The calculation formula of the normalized vegetation index is

\[
\text{NDVI} = \frac{\text{TM}_4 - \text{TM}_3}{\text{TM}_4 + \text{TM}_3}
\]

In addition, when collecting data, the sampling is guided according to the land use map obtained from the field survey. According to the analysis of land use maps and TM images, the ground features in the test area are divided into 8 categories. After sampling, a total of two sets of samples are obtained, each of which includes training samples and test samples.

4.2. Classification process
Basic SVM can only classify two types of problems, while remote sensing image is a multi-class classification problem, so multiple classifiers are constructed through a one-to-many method.

Based on the above, the remote sensing image classification process based on PCA-SVM is shown in Figure 2.

![Figure 2. PCA-SVM remote sensing image classification workflow](image-url)
Before training, normalize the original data to (-1, +1), and then use training samples to train various types of SVM classifiers. During training, first use secondary cross-validation to determine SVM parameters such as C and γ; Then use the test sample to verify the trained classifier, and select the appropriate model according to the classification accuracy and calculation time; finally use the selected SVM classifier to classify the entire test area. In order to compare the classification effects of SVM classifiers, this article also uses maximum likelihood classification (MLC) and neural network classification (BP neural network, BPNN) for classification.

4.3. Test results and analysis
From the perspective of classification accuracy, first of all, considering different classification methods, the overall accuracy of most SVM classifiers is much higher than the maximum likelihood method and neural network classification methods, which shows that SVM is a relatively stable and efficient classifier. It has good generalization ability; secondly, considering the different models of SVM, except the model based on Sigmoid kernel function and OAA multi-class method of sample set I (accuracy is 73.96%), the classification accuracy of other SVM models is not different Large, basically between 76%~78% (sample set 1) and 79%~82% (sample set 2). In comparison, the classification accuracy of the radial basis kernel function is slightly higher than that of the polynomial and Sigmoid kernel functions, at 77% and 81% respectively, and the classification accuracy of its various multi-class methods is also more stable than other kernels. The function is slightly better. Comparing the three multi-class methods, the classification accuracy of the OAO multi-class method is generally higher than that of the OAA method and the DAG method, especially the OAO classifier based on radial basis, its classification accuracy is the highest among all SVM classification models. Generally speaking, the stability of the OAA method is not as good as that of the OAO method and the DAG method. The reason is that the data of the two types of classifiers in the OAA method are not balanced. So in terms of accuracy and algorithm stability, the radial basis kernel function and OAO method are the best choices.

![Figure 3. The average missing points error of several algorithms](image)

In this paper, the remote sensing image classification result of the PCA-SVM algorithm has the highest recognition rate and the lowest average missing score error. The classification and recognition of remote sensing images has the lowest accuracy rate of SVM, mainly because the SVM algorithm has not undergone feature processing, the feature dimension is quite high, the classification time is long, the classification efficiency is the lowest, and the missing score is very serious. LBP-SVM is better than SVM, indicating that LBP can extract remote sensing images well, while PCA-SVM results are better than LBP-SVM. This is mainly because PCA effectively reduces the feature dimension, eliminates redundant information between features, and improves the image classification is correct, and the classification speed is accelerated. The correct classification of PCA? BPNN is lower than PCA-SVM. This is mainly because SVM has better anti-noise performance and generalization ability.
than BPNN. Comparative experimental results show that PCA-SVM is an effective and high accuracy rate. The remote sensing image classification method can well realize the automatic classification of remote sensing images.

5. Conclusion
This paper uses artificial intelligence algorithms to classify multi-source remote sensing images, and compares various supporting satellite machine models from the perspectives of classification accuracy, calculation time, and algorithm stability. Among them, the choice of models includes kernel function and multi-class method select. The results show that the algorithm in this paper has high classification accuracy, and remote sensing data plus other auxiliary data can improve the classification accuracy.

References
[1] Tang H, Mao L, Zeng S, et al. Discriminative dictionary learning algorithm with pairwise local constraints for histopathological image classification [J]. Medical & Biological Engineering & Computing, 2021, 59(6):1-8.
[2] Calisto F M, Santiago C, Nunes N, et al. Introduction of Human-Centric AI Assistant to Aid Radiologists for Multimodal Breast Image Classification [J]. International Journal of Human-Computer Studies, 2021, 150(3):102-116.
[3] Gatos I, Tsantis S, Spiliopoulos S, et al. A new automated quantification algorithm for the detection and evaluation of focal liver lesions with contrast-enhanced ultrasound [J]. Medical Physics, 2015, 42(7):3948-3957.
[4] Novoa R A, Olivier G, Ko J M. Marking the Path Toward Artificial Intelligence-Based Image Classification in Dermatology [J]. JAMA Dermatology, 2019, 155(10):1-7.
[5] Jitaree S, Phinyomark A, Boonyaphiphat P, et al. Cell type classifiers for breast cancer microscopic images based on fractal dimension texture analysis of image color layers [J]. Scanning, 2015, 37(2):145-151.
[6] Rigon A, Soda P, Infantino M, et al. FRI0574Slim-System in ANA IIF Reading: A Multicentre Study for the Assessment of Interobserver Reading Variability [J]. Annals of the Rheumatic Diseases, 2015, 74(2):632-636.
[7] Bousfiha A, Jeddane L, Picard C, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies [J]. Journal of Clinical Immunology, 2017, 38(1):59-62.
[8] Surai, P, F, et al. Glutathione peroxidases in poultry biology: Part 1. Classification and mechanisms of action [J]. World’s Poultry Science journal, 2018, 74(2):185-197.
[9] Kumar S, Gezari S, Heinis S, et al. Selection of Burst-like Transients and Stochastic Variables Using Multi-Band Image Differentiating in the Pan-STARRS1 Medium-Deep Survey [J]. Astrophysical Journal, 2015, 802(1):56-67.
[10] Xu J, Feng G, Zhao T, et al. Remote sensing image classification based on semi-supervised adaptive interval type-2 fuzzy c-means algorithm [J]. Computers & geosciences, 2019, 131(Oct.):132-143.