Hyperspectral Image Classification Based on GS-CatBoost Model

Yan Wang, Xiaoli Sun* and Guochen Li
School of Computer and Communications, Lanzhou University of Technology, 287 Lan-gong-ping Road, Lanzhou, Gansu 730050, China
*Corresponding Author Email: 3189477072@qq.com

Abstract. Aiming at the features of hyperspectral images with large amount of redundant information of hyperspectral images, small correlation between features and categories, small sample size, and high-dimensionality data, a model of hyperspectral image classification based on an improved mRMR band selection method and GS-CatBoost is proposed. First of all, using the mRMR improved by K-L to select the optimal band subset with high correlation between features and category, low redundancy, and large amount of information. Then, the Gabor filter is embedded in PCA and LDA combined dimensionality reduction to complete the extraction of space-spectrum joint features. Finally, using grid search to optimize the CatBoost algorithm to classify the extracted features. Experiments show that the proposed model fully and effectively extracts the space-spectrum joint features of hyperspectral images. In the Salinas Scene dataset, OA is 97.87%, AA is 99.01%, and Kappa coefficient is 0.9763.

1. Introduction
Hyperspectral image [1-2] is a three-dimensional data cube, which can provide rich spatial and spectral information, so the combined features of space spectrum can be used for effective classification. The ground object classification of hyperspectral images is widely adapted. It can be used for precision agricultural monitoring in ecology, Pollutant types and concentration indicators analyzing in environmental monitoring, and smoke and dust distribution and coal spontaneous combustion information analyzing in fire cause analysis.

Hyperspectral images have the phenomenon of large amount of redundant information, strong correlation, and high dimensionality. Before performing hyperspectral images classification, it is necessary to select bands [3-5] or reduce the data dimensionality reduction to complete the feature extraction [6-11]. Han et al. [12] proposed a method of band selection using sub-pixel target detection to achieve dimensionality reduction and feature extraction of hyperspectral images. Feng et al. [13] proposed to filter bands with maximum correlation and minimum redundancy criteria to solve the problem of the removal of redundant information and the correlation between features and categories. Wang et al. [14] proposed a combination of PCA and wavelet transform to solve the problem of extracting multi-scale principal component features. Heng-Chao et al. [15] proposed to use the compound kernel to project the Gabor feature of the hyperspectral image onto the new classifier, which solved the problem of feature extraction and recognition enhancement from different angles. Liu et al. [16] proposed two improved ensemble learning[17] with different cores of SVM in the AdaBoost framework to solve the problem of improving the classification accuracy of hyperspectral image spatial spectrum joint features. Although researchers have proposed a large number of hyperspectral image classification methods [18-19], these methods still have the problems of loss of some effective information caused by dimensionality reduction and insufficient feature extraction.
In conclusion, this paper aims at the problems of hyperspectral images, and proposes a classification model based on improved Max-Relevance and Min-Redundancy (mRMR) band selection and Grid Search-CatBoost (GS-CatBoost). First, the optimal band subset is selected by band selection, and then space-spectrum joint feature extraction is performed by combining dimensionality reduction with embedded Gabor features, and finally improving classification accuracy with the model of CatBoost optimized by grid search.

2. Improved Band Selection
To solve the problems of low correlation between hyperspectral image features and categories, large data redundancy, and loss of some information caused by dimensionality reduction, a band selection method of mRMR improved by K-L divergence is proposed. Mutual information is the sum of the information of two continuous variables, but the data features and categories are discrete data, so K-L divergence is used as the objective function of mRMR. The improved mRMR algorithm can realize the maximum correlation between data features and categories, but it cannot eliminate the redundant information between the bands of hyperspectral image. Therefore, the K-L divergence and the mRMR improved by the K-L divergence are used to select the optimal band subset, and solve the problems of small category correlation and large redundancy.

The K-L divergence is used to calculate the information difference between any two bands. The larger the information difference, the smaller the correlation. After sorting in descending order, the bands are searched in order to filter out the optimal band subsets with mass information and low correlation.

K-L divergence formula is shown in (1):

$$D_{KL}(x_i \parallel x_j) = \sum_{i,j}^n x_{it} \log \frac{x_{it}}{x_{jt}}$$  \hspace{1cm} (1)

Where n is the number of pixels in each band, xi is the i-th band, and xit is the t-th spectral feature value of the i-th band.

The improved mRMR is used to calculate the information difference between features and categories. The larger the information difference, the smaller the correlation. After sorting in ascending order, sequential search is used to perform secondary screening of the optimal band subset to ensure that the band subset has the largest amount of original information, the smallest relevance, and enhanced category discrimination.

mRMR is as shown in the formula (2):

$$\max D(S, c), D = \frac{1}{|S|} \sum_{i \in S} I(x_i; c)$$  \hspace{1cm} (2)

The improved mRMR is shown in formula (3):

$$\min F(S, c), F = \frac{1}{|S|} \sum_{i \in S} D_{KL}(x_i \parallel c)$$  \hspace{1cm} (3)

Among them, S is a band subset, |S| is the number of bands, c is the number of categories, and I(xi;c) and D_{KL}(x||c) is the mutual information between the i-th band and the ground object category c.

3. Optimized CatBoost Model
In order to further improve the classification accuracy of data features after band selection and combination dimensionality reduction preprocessing, a classification model of GS-CatBoost is constructed to solve the problems of the instability of the randomly selected parameter model and the time increasing for model training with the increased numbers of combined parameters. Aiming at the problem of low classification accuracy of manual adjustment of CatBoost parameters and low model scores caused by multiple debugging experiments, this article uses grid search to optimize the CatBoost algorithm. Grid Search (GS) is a parameter optimization algorithm for exhaustive search. Firstly, we use coordinate descent for rapid tuning and the idea of grid search algorithm, learning rate,
max_depth, n_estimators, l2_leaf_reg, loss_function and other parameters are optimized through loop traversal cross validation to optimize the CatBoost classification model, improve the classification accuracy, and strengthen the stability of the model. Reduce classification time.

The specific steps of this model are as follows:

1) Obtain hyperspectral image data and divide the data set into a training set and a testing set.

Equation (1) was used to sequentially search the original hyperspectral image to filter out the band subset with small correlation and large amount of information; equation (3) was used to perform a sequential search on the band subset with the largest amount of information to filter out the optimal band subset with low correlation between bands and high correlation between features and categories;

2) The key components of the optimal band subset is Calculated by PCA; use LDA to reduce the intra-class spacing and increase the inter-class spacing.

3) Perform linear discriminant analysis on the Gabor spatial features of 5 scales and 8 directions for the first c principal components, and complete the extraction of space-spectrum joint features;

4) The grid search algorithm is used in CatBoost to construct the GS-CatBoost classification model, and the training set is used to train the GS-CatBoost model.

5) Then we use the testing set to test the GS-CatBoost classification model, and use the verification set to verify the GS-CatBoost classification model.

4. Experimental Results and Analysis

4.1 Data Set Selection

The public data set Salinas scene is an image of Salinas Valley, California, USA taken by the AVIRIS imaging spectrometer. The data has a spatial resolution of 3.7m and includes 16 types of ground features, a total of 512*217 Pixels, and 224 bands. After removing the 20 noise bands, 204 bands can now be used. Figure 1 shows the false-color image and ground reference map of the Salinas scene dataset.

The public data set Indian Pines is an imaging of the Indiana agricultural and forestry hyperspectral experimental site in northwestern Indiana, USA taken by the AVIRIS in 1992. The data has a spatial resolution of 20m and includes 16 ground features. There are 145*145 pixels and 220 bands in total. After removing 20 noise bands, 200 bands can be used. Figure 1 shows the specific categories and numbers of the Indian Pines data set.

![Salinas scene false color image.](a)

![Salinas scene ground reference map.](b)

![Indian Pines false color image.](c)

![Indian Pines ground reference map.](d)

Figure 1. The false-color image and ground reference map of the dataset.

4.2 Experimental Environment And Index Selection

4.2.1 Experimental environment

The experimental environment of this article is a PC with Intel(R) Celeron(R) 3205U clocked at 1.50GHz and 8GB of memory. The specific program is written by Pycharm2017.3.6.

4.2.2 Indicator selection

This model selects the public data sets Salinas scene and Indian Pines for experiments. During the process of training, testing and verification, we randomly select 60% of each
type of ground object as the training set and 40% as the test set. Repeat the experiment 20 times and take the average value as the final experimental result. In this paper, the overall classification accuracy (OA), the average recall rate of each class (AA) and Kappa coefficient are used to evaluate the classification performance.

4.3 Band Selection Parameter Comparison Experiment

In this section, the Salinas scene data set is used to carry out the image preprocessing part parameter setting experiment. It is the comparison experiment of band parameters selected by the improved mRMR band. Where $W_kl$ is the number of K-L divergence bands, $W_i$ is the number of improved mRMR bands, $P$ is the number of PCA dimensions, $L$ is the number of LDA dimensions, $K$ is the scale, and $\lambda$ is the wavelength.

It can be known from Table 1 that $P=15$, $L=15$, and $\lambda=5$ are sets according to experimental debugging, where $W_kl-W_i$ is abbreviated as the number of bands. Among them, when $W_kl=150$ and $W_i =120$ reach the highest, the coefficients of OA, AA, and Kappa increase by 1.44%, 0.71% and 0.016, respectively.

| data set       | Band   | OA     | AA     | Kappa |
|----------------|--------|--------|--------|-------|
| Salinas scene  | 160-140| 97.68% | 98.88% | 0.9742|
|                | 160-130| 97.86% | 98.96% | 0.9761|
|                | 160-120| 97.86% | 98.97% | 0.9762|
|                | 150-120| 97.87% | 99.01% | 0.9763|
|                | 150-130| 96.48% | 98.33% | 0.9608|
|                | 150-140| 96.43% | 98.30% | 0.9603|

4.4 Analysis of Model Experiment Results

This model selects the Salinas scene data set and Indian Pines data set for experiments. In order to ensure the objectivity and fairness of the experiment, all of the following experiments use the same experimental conditions and the data preprocessing is set as the optimal parameter of this model.

4.4.1 Comparison of different band selection models

In this paper K-L and mRMR (KM) is compared with the model of K-L and mRMR for band selection to verify the correctness of the improved mRMR band selection model proposed in this section. From Table 2 on the Salinas scene data set, KM has increased by 2.38% than KL in the OA, and increased by 2.44% than mRMR in the OA. In the Indian Pines data set, KM increased by 4.22% than KL in the OA, and increased by 2.44% than mRMR in the OA.
Table 2. Classification results of different band selection models

| Methods       | KL     | mRMR    | KM     |
|---------------|--------|---------|--------|
| Salinas scene | OA     | 95.49%  | 95.26% | 97.87% |
|               | AA     | 97.73%  | 97.81% | 99.01% |
|               | Kappa  | 0.9498  | 0.9472 | 0.9763 |
| Indian Pines  | OA     | 91.00%  | 89.78% | 95.22% |
|               | AA     | 90.86%  | 87.77% | 93.14% |
|               | Kappa  | 0.8972  | 0.8832 | 0.9454 |

4.4.2 Analysis of model experiment result In this section, K-L-mRMR- GSCatBoost (GS KM-GSCatBoost) and PCA-CatBoost (P-CatBoost), LDA-CatBoost (L-CatBoost), PCA-Gabor-CatBoost (PG-CatBoost), PCA-Gabor-LDA-CatBoost (PGL-CatBoost) are compared to verify the correctness of model feature extraction. It can be seen from Table 3 that the Salinas scene dataset that KM-GSCatBoost is 1.4%, 0.69%, and 0.156% higher than the highest PGL-CatBoost OA, AA, and Kappa respectively; The classification accuracy of this model is higher than that of other feature extraction methods. On the Indian Pines dataset, the OA of KM-GSCatBoost is 2.64%, 0.65% and 2.33% higher than that of P-CatBoost, L-CatBoost, and PGL-CatBoost, respectively, but the classification performance is lower than that of PG-CatBoost.

Table 3. Classification results of different feature extraction model

| Methods       | P-CatBoost | L-CatBoost | PG-CatBoost | PGL-CatBoost | KM-GSCatBoost |
|---------------|------------|------------|-------------|--------------|---------------|
| Salinas scene | OA         | 95.12%     | 94.69%      | 95.82%       | 96.47%        | 97.87%        |
|               | AA         | 97.49%     | 97.45%      | 97.86%       | 98.32%        | 99.01%        |
|               | Kappa      | 0.9456     | 0.9409      | 0.9534       | 0.9607        | 0.9763        |
| Indian Pines  | OA         | 92.58%     | 94.57%      | 96.7%        | 92.89%        | 95.22%        |
|               | AA         | 91.57%     | 94.94%      | 95.6%        | 92.92%        | 93.14%        |
|               | Kappa      | 0.9153     | 0.938       | 0.9624       | 0.9188        | 0.9454        |

Figure 2. The classification performance of different feature extraction models in the Salinas scene dataset.
Figure 3. The classification performance of different feature extraction models in the Indian Pines dataset.

In the figure, a-f are the classification diagrams of different feature extraction models in the Salinas scene dataset and the Indian Pines dataset (the order is the same as in Table 3). In Figure 2, the noise of KM-GSCatBoost on the Salinas scene data set shows a significant reduction. In Figure 3, the noise of KM-GSCatBoost on the Indian Pines data set is the least, Wood's area classification is almost completely correct. It can be seen that the extraction and classification effect of the space-spectrum joint feature of this model is better, and the model has a strong generalization ability.

4.4.3 Comparison of different ways to optimize CatBoost hyperparameter models All CatBoost parameters have an important impact on accuracy, and the classification accuracy can be improved when the parameters are optimal. Set learning_rate=0.03, loss_function=Multiclass, parameter depth and n_estimators are randomly set to 3 and 200 (d3-n200 for short), 4 and 1000, 6 and 400. The comparison of grid search optimization CatBoost and manual tuning parameters is shown in Table 4. It can be seen from table 4 that GS-CatBoost is significantly better than random tuning.

| Parameter debugging | d3-n200 | d4-n1000 | d6-n400 | GS       |
|---------------------|---------|----------|---------|----------|
| OA                  | 95.48%  | 96.48%   | 97.38%  | 97.87%   |
| AA                  | 97.66%  | 98.33%   | 98.69%  | 99.01%   |
| Kappa               | 0.9496  | 0.9608   | 0.9708  | 0.9763   |

4.4.4 Comparison of different classifier models In this section, the model with CatBoost as the classifier is compared with the SVM, Adaboost, XGboost, and LightGBM classifiers. It can be seen from Table 5 that in the Salinas scene data set, the OA of this model is 2.63% higher than SVM respectively, the OA of this model is 2.05% higher than XGboost respectively. In the Indian Pines data set, the OA of this model is 9.31% higher than SVM respectively. The OA of this model is 2.51% higher than LightGBM respectively. It shows that the boosting algorithm performs much higher than the traditional machine learning algorithm SVM, and CatBoost has the best classification.
Table 5. Classification results of different classifier models

| Methods   | PCA-SVM | PCA-Adaboost | PCA-XGboost | PCA-LightGBM | KM-GSCatBoost |
|-----------|---------|--------------|-------------|--------------|--------------|
| Salinas scene | OA      | 95.24%       | 88.84%      | 95.82%       | 94.11%       | 97.87%       |
|           | AA      | 97.91%       | 95.72%      | 97.86%       | 97.31%       | 99.01%       |
|           | Kappa   | 0.947        | 0.8759      | 0.9534       | 0.9343       | 0.9763       |
| Indian Pines | OA      | 85.91%       | 92.71%      | 89.66%       | 91.47%       | 95.22%       |
|           | AA      | 88.19%       | 91.08%      | 88.7%        | 90.24%       | 93.14%       |
|           | Kappa   | 0.8387       | 0.9165      | 0.8817       | 0.9024       | 0.9454       |

a-f are the classification results of different classifier models in the Salinas scene dataset and the Indian Pines dataset (the order is the same as in Table 8). In Figure 5, the noise of the KM-GSCatBoost model classification result in the Salinas scene data set is greatly reduced. Among them, It can be seen from Figure 6 that the classification result of this model on the Indian Pines dataset has the least noise and the area of the noise is greatly reduced, but the overall noise is obvious. Among them, Woods classification is almost completely correct. From this, it can be seen that the classification model in this article has achieved good classification results and has strong model generalization ability.

5. Conclusion
This article proposes new improved methods for band selection, dimensionality reduction and classification methods. The improved method realizes the dimensionality reduction of hyperspectral
images and preserves the original information amount and category-related information to the greatest extent. Improved mRMR band selection is used to improve category relevance, combined dimensionality reduction and Gabor filtering to extract space spectrum joint features, solve the problem of high data dimensions and small samples, and using GS-CatBoost to improve classification accuracy. Experiments show that the model proposed in this paper not only realizes the extraction of space-spectrum joint features in hyperspectral image classification, but also greatly improves the classification accuracy. However, there are still the following limitations: the existing search method is improved to select the better band subset, the feature extraction method is improved and the feature fusion algorithm is added to solve the problem of insufficient feature extraction, the classification accuracy is further improve.

6. References
[1] Binge C, Xiaoyun X, Siyuan H , Jiandi C and Yan L 2018 . J. Remote Sensing Semi-supervised classification of hyperspectral images based on extended label propagation and rolling guidance filtering 10(4) p 515.
[2] Mughees A and Tao L (2019) J. Tsinghua University and Technology Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images 24(02) p 63-74.
[3] Zhang W, Li X, Dou Y, and Zhao L (2018) J. Journal of Applied Remote Sensing Fast linear-prediction-based band selection method for hyperspectral image analysis 12(1) p 1.
[4] Barman B, and Patra S (2019) IET Image Processing An empirical study of neighborhood rough sets based band selection techniques for classification of hyperspectral images 13(8) pp 1266-1279.
[5] Zachary T, Dar R, SanderV, Casas ángeles, Carlos R and Susan U (2018) J. Remote Sensing Evaluating endmember and band selection techniques for multiple endmember spectral mixture analysis using post-fire imaging spectroscopy 10(3) p 389.
[6] Hu Y, Zhang J, Ma Y, Li X, Sun Q and An J. J 2019 J. Acta Oceanologica Sinica Deep learning classification of coastal wetland hyperspectral image combined spectra and texture features: a case study of huanghe(yellow) river estuary wetland 38(05) pp 142-150.
[7] Mughees A, and Tao L (2019) J. Tsinghua University and Technology Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images 24(02) pp 63-74.
[8] Garcia-Salgado B. P, Ponomaryov V. I, Sadovnychiy S and Reyes-Reyes R. J 2020 J. Journal of Applied Remote Sensing Efficient dimension reduction of hyperspectral images for big data remote sensing applications 14(3) p 1.
[9] Ding L, Fang W, Luo H, Love P. E. D, Zhong B, and Ouyang X (2018) J. Automation in Construction A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory 86 p124.
[10] Uddin M. P, , Al Mamun M. , and Hossain M. A. (2019) J. International Journal of Remote Sensing Effective feature extraction through segmentation-based folded-pca for hyperspectral image classification 40(17-18) pp 1-31.
[11] Akbari DJ 2020. J. Arabian Journal of Geosciences A novel method for spectral-spatial classification of hyperspectral images with a high spatial resolution 13(23) pp 1-10.
[12] Han S, Kerekes J, Higbee S, Siegel L, and Pertica A (2019) J. Applied Optics Band selection method for subpixel target detection using only the target reflectance signature 58(11) p 2981.
[13] Feng J, Licheng J, Xiangrong S, and Tao, et al. (2016) J. Pattern Recognition: The Journal of the Pattern Recognition Society Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images 51 pp 295-309
[14] Wang G, Zhang Y, Liu C, Xie Q, and Xu Y (2019) J. Journal of Intelligent Manufacturing A new tool wear monitoring method based on multi-scale pca 30(1) pp 113-122.
[15] Heng-Chao L, Hong-Lian Z, Lei P, and Qian D (2018) J. Electronics Letters Gabor feature-based composite kernel method for hyperspectral image classification 54(10) pp 628-630.
[16] Huang G, Wu L, Ma X, Zhang W, Fan J, and Yu X, et al. (2019) J. Journal of
Hydrology Evaluation of catboost method for prediction of reference evapotranspiration in humid regions 574 pp 1029-1041.

[17] Yan Y (2020) JIOP Conference Series Materials ence and Engineering Application of data mining method based on integrated learning in early warning research of smes' financial crisis 740 p 012200.

[18] Liu X, Chuai G, Gao W, and Zhang K (2018) J. Eurasip Journal on Wireless Communications & Networking Ga-adaboostsvm classifier empowered wireless network diagnosis 2018(1) p 77.

[19] Zhu X, and Li G. J 2019 J. International Journal of Food Properties Rapid detection and visualization of slight bruise on apples using hyperspectral imaging 22(1) pp 1709-1719.