Simulated Annealing-Based Optimization for Band Selection in Hyperspectral Image Classification

Said Khelifa[1], Fatima Boukhatem[2], Leila Benaissa Kaddar[3]

[1] University of Science and Technology Mohamed Boudiaf,
SIMPA Laboratory,
Algeria

[2] University of Sidi Bel Abbes Djillali Liabes,
Algeria

[3] University of Mustapha Stambouli Mascara,
Algeria

said.khelifa@univ-usto.dz, fatima.boukhatem@univ-sba.dz,
l.benaissa@univ-mascara.dz

Abstract. In this paper, a new optimization based framework for hyperspectral image classification problem is proposed. Band selection is a primordial step in supervised/unsupervised hyperspectral image classification. It attempts to select an optimal subset of spectral bands from the entire set of hyperspectral cube. This subset is considered as the relevant informative subset of bands. The advantage of an efficient band selection approach is to reduce the hughes phenomenon by removing irrelevant and redundant bands. In this study, we propose a new objective function for the band selection problem by using Simulated Annealing as an optimization method. The proposed approach is tested on three Hyperspectral Images largely used in the literature. Experimental results show the performance and efficiency of the proposed approach.

Keywords. Optimization, band selection, classification, bagging, correlation, simulated annealing.

1 Introduction

Nowadays, feature selection has become a very active research field in machine learning. Used in supervised and unsupervised learning, the feature selection process has been an indispensable tool to reduce the dimensionality of datasets and consequently reduce the hughes phenomenon.

This last is detected when a large number of features are present compared to a small number of samples. This is why, feature selection is largely used in many domains such as gene selection in DNA microarray datasets, band selection in hyperspectral images, feature selection in image classification and recognition [13, 1, 7].

In supervised classification, it is crucial to select features carefully in order to achieve high classification accuracy and constructing a good classification model. It attempts to select the relevant and informative features to reduce the computational complexity [2, 12, 4].

The feature selection approach is divided into three classes: wrapper, embedded and filter. Many feature selection approaches have been proposed in various applications.

In [10], Medjahed et al. have proposed a new feature selection approach based on SVM-RFE and Binary Dragon fly. The objective function is a combination of three classifiers (SVM-LS, C-SVM and v-SVM).

Their work was applied to the DNA micro array dataset. Gold et al. [5] proposed a Bayesian viewpoint of SVM classifiers to determine irrelevant features by adjusting parameter values.
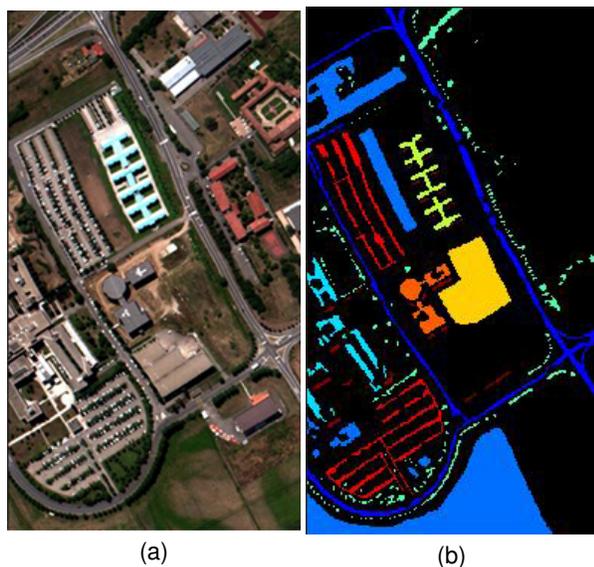
Fig. 1. Pavia University hyperspectral image. (a) RGB Image. (b) Ground truth
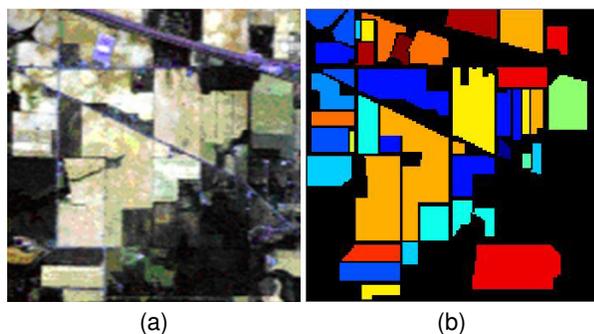

Fig. 2. Indian Pines hyperspectral image. (a) Color image. (b) Ground truth

In [6], performance is measured through the accuracy rate of the classifier. Medjahed et al. [8] have proposed two approaches based on the Threshold Accepting and the Microcanonical Annealing used for feature selection.

Several medical datasets have been used to test the approaches that have shown their effectiveness through experimental results.

In [9] , the authors developed a new objective function that contains two important terms: the first is the classification accuracy rate, and the second is the Jeffries Matusita distance.

The proposed objective function is optimized using the Grey Wolf Optimizer Algorithm. Their work aims to select the relevant bands in hyperspectral images that provided a high classification accuracy rate and make sure the classes are well separated. Minimum Noise Band Selection is a new approach for selecting bands proposed in [14]. Based on low correlation and high SNR , this approach determines the quality of each band. Search efficiency can be improved by sequential backward selection.

A new band selection framework has been developed in [15]. By using an evolutionary strategy associated with groupwise selection strategies, the proposed approach reduces the computational burden.

Progressive Band Selection (PBS), a new method of selecting bands, has been developed In [3]. PBS differs from traditional band selection. As opposed to the former, which adapts the number of selected bands, p, to the various end members used for spectral unmixing, the latter fixes, p, at the same constant value for all end members.

In this paper, a new approach to the band selection problem is proposed. The proposed approach is based on Simulated Annealing which is a very used metaheuristic in the literature, and we propose a new objective function composed of two terms.

Three hyperspectral images widely used for testing image classification, namely are : Salinas Scene, Pavia University Scene and Indian Pines Scene are used to evaluate the performance of the proposed approach.

The remainder of this paper is structured as follows: Section 2 describes the proposed approach, Section 3 describes and discusses the experimental results and in Section 4 the conclusion and some perspectives are drawn.

2 The Proposed Approach

Typically, each feature selection approach has an objective function, a search strategy, a generator of subset candidates, and a stop criterion.

In our study, the band selection problem has the following mathematical form. Let's:

$$D = \{(x^1,\ y_1),\ \ldots, \{x^n,\ y_n\}\}, \qquad (1)$$

where $x^j = (x_1^j, .., x_N^j)$ is the j-*th* instances, y is the classes and $X = \{X_1, X_2, \ldots, X_N\}$ is the set of features. Any feature selection approach aims to select a subset of X which is considered as the optimal subset.

2.1 The Objective Function

In feature selection, the objective function computes a certain measure for each generated subset. It can have many forms. Generally, we can have a filter function if it used a statistical measure independently of the classifier system, and it can be a wrapper function if it is computed from a classifier system.

In this study, we propose a new objective function based on two important terms. The first term is based on bagging approach.

The advantage of bagging is to reduce the effect of overfitting and improves classification and generalization. The basic idea is to generate several decision trees under a different subset of predictors and combines the results of these decision trees. The proposed score is computed as follows:

$$R(X) = 1 - \frac{1}{n} \sum_{j=1}^{n} \text{fct\_class}(\text{classify}(x^j)_{\text{Tree}_{1,\ldots,n}}), \quad (2)$$

$$\text{classify}(x^j)_{\text{Tree}_i} = \begin{cases} 1, & \text{if correctly classify } x^j, \\ 0, & \text{elsewise}, \end{cases} \quad (3)$$

where, $R(X)$ is the score computed over the selected features ($X_i = 1$). $\text{classify}(x^j)_{\text{Tree}_i}$, is equal to 1 if all the decision trees ($\text{Tree}_i, i = 1, \ldots, nb$) has the same values. $\text{classify}(x^j)_{\text{Tree}_1}$ is the class of the instance x^j. n is the number of instances used for test.

In other terms, $R(X)$ computes the rate of misclassified instances over all the decision trees. The second term of the objective function is the correlation between the features. This term is given as follows:

$$C(X) = \frac{1}{N \times (N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} C_p(X_i, X_j), \quad (4)$$

Algorithm 1 Simulated annealing for band selection

1: Read the training set x.

2: Initialize the system by selecting randomly a subset of bands x_k;

3: Calculate the objective function $J(x_k)$ by calculating $R(x_k)$ and $C(x_k)$.

4: Choose another subset of bands x_{k+1}.

5: Calculate the objective function $J(x_{k+1})$ by calculating $R(x_{k+1})$ and $C(x_{k+1})$.

6: **if** $J(x_{k+1}) \leq J(x_k)$ **then**

7:     accept transition to the new state $x_k = x_{k+1}$

8: **else**

9:     accept transition with probability $\exp\left(-\frac{\Delta J}{T}\right)$

10: **end if**

11: Decreasing the temperature by a cooling schedule.

12: Repeat steps until the end.

13: Save the best subset of bands x_{best}.

---

where, $C_p(X_i, X_j)$ is the Pearson correlation coefficient computed between the features X_i and X_j. $C(X)$ is the correlation computed between all the selected features ($X_i = 1$). N is the number of selected features. The final form of the objective function is given as follows:

$$J(X) = \alpha \times R(X) + \beta \times C(X), \qquad (5)$$

where, α, β are the weight coefficients. Minimizing the objective function $J(X)$ is the main goal.

2.2 Simulated Annealing Algorithm

The second important criterion in the feature selection process is the search strategy. In our approach, we propose to use the Simulated Annealing Algorithm which is a meta heuristic widely used in image classification and it showed its performance. Simulated Annealing is an optimization approach for approximating the global optimum in a large search space.

The origin of this algorithm is the annealing in metallurgy. The process is to heat the metal to high temperature and cooled slowly.
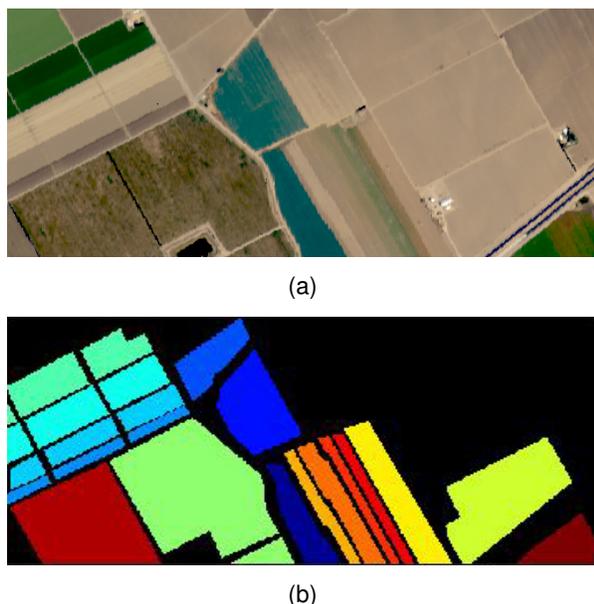
(a)

(b)

Fig. 3. Salinas hyperspectral image. (a) Color compose. (b) Ground truth

Table 1. Average accuracy, overall accuracy and individual classaccuracy obtained by the proposed approach in each hyperspectral image

Class	Hyperspectral Images		
	Pavia University	Indian Pines	Salinas
1	87,16	57,14	99,25
2	90,03	60,09	99,64
3	66,75	53,21	96,54
4	86,68	39,86	99,16
5	98,64	82,41	97,82
6	70,01	82,42	99,37
7	77,07	82,35	99,30
8	77,87	87,46	76,16
9	100,00	41,67	98,60
10		54,45	89,37
11		68,57	95,48
12		46,35	98,01
13		82,11	98,00
14		88,54	92,52
15		45,26	69,14
16		89,29	96,50
AA	83,80	66,32	94,06
OA	84,89	67,35	89,23

The first step of the algorithm is to set the system to an arbitrary solution and by using a sequence of transitions, we achieve the final state which is considered as the optimal solution:

$$P\left(x_{k+1}|x_k\right) = \min\left[1, \ \exp\left(-\frac{E\left(x_{k+1}\right) - E\left(x_k\right)}{T}\right)\right], \quad (6)$$

where $E\left(x_k\right)$ is the energy function value at the kth iteration. Using the Boltzmann distribution at temperature T. Metropolis et al. [11] proposed an algorithm to simulate the system's behavior.

The algorithm schema can be as shown in Algorithm 1. Note there are some parameters in the simulated annealing algorithm which should be set carefully which are: initial temperature, cooling schema, stop criterion.

3 Experimental Results

The results we obtained by applying our approach are presented based on three hyperspectral images. We evaluate the proposed approach in terms of overall accuracy rate, average accuracy rate and individual class accuracy rate.

3.1 Dataset

Pavia University image is taken ROSIS over the Pavia, Northern, Italy. It was composed of 103 bands in the spectral range 0.4 to $0.86\,\mu\text{m}$. the image size is 610×340.

A total of 9 classes are contained in this image: Asphalt, Gravel, Meadows, Bare Soil, Trees, PaintedMetal Sheets, Bitumen, Shadows, and Self-Blocking Bricks. Figure 1 shows the Pavia University RGB color and the Ground truth.

Indian Pines image was taken by AVIRIS over the Indian Pines, North-Western Indiana, USA. It consists of 145×145 pixels and 220 spectral bands in the spectral range 0.5 to $2.5\,\mu\text{m}$.

The ground truth was composed of 16 classes, namely: Alfalfa, Corn, Corn-notill, Corn-mintill, Grass-pasture, Grass-trees, Grass-pasture-mowed, Hay-windrowed, Oats, Soybean-notill, Soybean-mintill,Soybean-clean, Wheat, Woods, Buildings-Grass-Trees-Drives, and Stone-Steel-Towers. Figure 2 shows the Indian Pines RGB color and the Ground truth.

Salinas hyperspectral image is taken by AVIRIS over Salinas Valley, California, USA. The area covered 512×217 in a spectral range of 0.4 to 2.5 μm.

This image is composed of 224 bands and 16 ground truth classes: Fallow, Fallow-smooth, Fallow-rough-plow, Broccoli-green-weeds-1, Corn-senesced-green-weeds, Broccoli-green-weeds-2, Celery, Stubble, Grapes-untrained, Soil-vinyard-develop, Lettuce-romaine-7wk, Lettuce-romaine-6wk, Lettuce-romaine-5wk, Lettuce-romaine-4wk, Vineyard-vertical-trellis, and Vineyard-untrained. Figure 3 shows the Salinas RGB color and the Ground truth.

3.2 Results and Discussion

In this experimentation, we propose to split the data into three steps, Training, Testing and Validation. The Simulated annealing is set as follows: initial temperature is set to 1000. We use a geometric cooling schema with 0.98 as a parameter of cooling. Table 1 shows the results obtained by the proposed approach.

Table 1 illustrated the results obtained by the proposed approach using three widely used hyperspectral images : Salinas, Pavia University and Indian Pines.

The first column shows the class number, the second column represents the Pavia University results, the third column represents the Indian Pines results and the last column represents the Salinas results.

The experimental results are conducted in terms of average accuracy (AA), overall accuracy (OA) and individual class accuracy. The two last columns of the table represent respectively the AA and OA.

The algorithm is also running 100 times under different training, testing and validation sets randomly selected in each iteration.

Until 10% of the total number of pixels are used for training, the remaining pixels are used for testing and validation phases. The analysis of the results described in table 1 demonstrates the performance of the proposed approach.
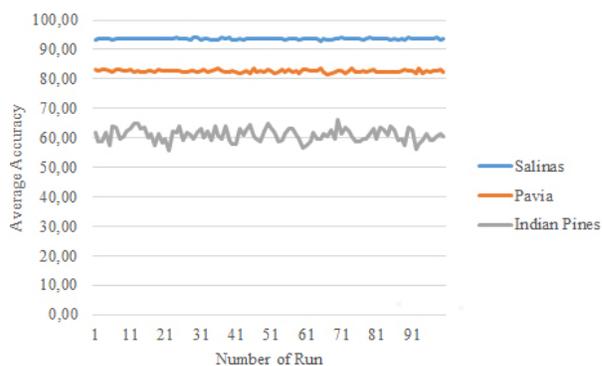

Fig. 4. Average Accuracy obtained by the proposed approach for each hyperspectral image
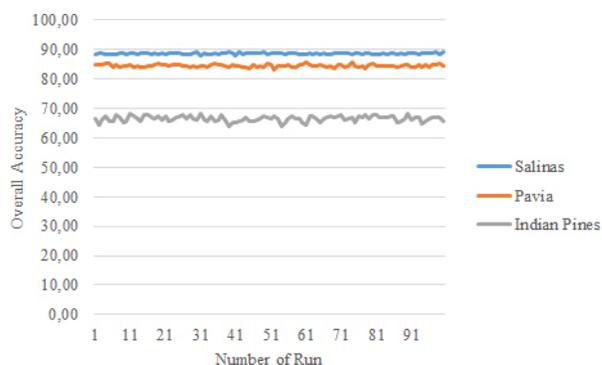

Fig. 5. Overall Accuracy obtained by the proposed approach for each hyperspectral image

The classification accuracy rate obtained is very high with an advantage to Salinas images which produced $97,06\%$ of average accuracy and $89,23\%$ of overall accuracy.

Figures 4 and 5 illustrate the AA and OA obtained in each execution. Figure 6 shows the classification map for each image, produced by the proposed approach.

We compare the proposed approach to three optimization method by using the same objective function. These methods are: Tabu search (TS), Threshold accepting (TA) and Genetic Algorithm (GA).

The comparaison has been done on Pavia University Image. Table 2 describes the results obtained by the approaches. We conduct the comparison by using the same objective function under the Pavia University Image.
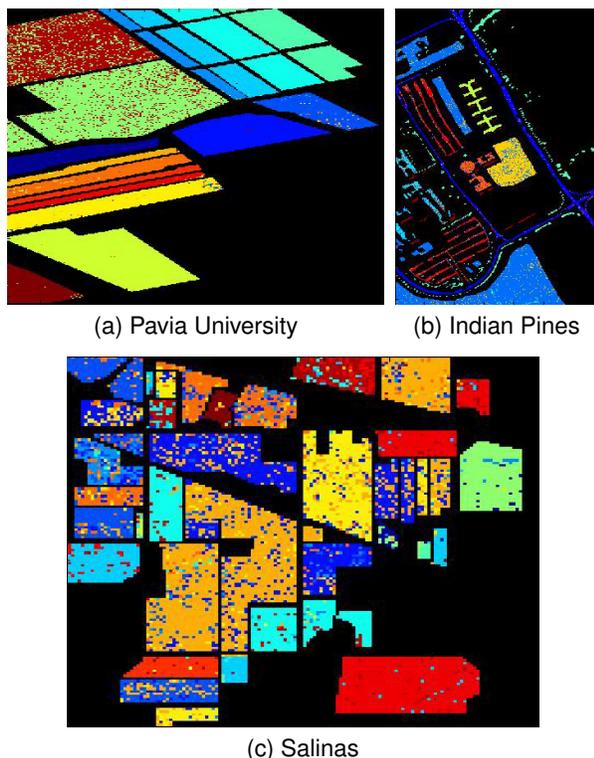
(a) Pavia University   (b) Indian Pines

(c) Salinas

Fig. 6. The classification maps obtained by the proposed approach for Salinas, Pavia University, and Indian Pines hyperspectral images

Table 2. Average accuracy obtained by the proposed approach and compared to other approaches

Method	AA	OA
This study	83,80	84,89
TS	70.53	75.41
TA	74.86	76.20
GA	83.61	84.02

The analysis of the results shows that the values obtained by the proposed approach are slightly better than those of (TS) and (TA). Compare to the results obtained by (GA), they are almost the same.

4 Conclusion

In this paper, we present a new objective function for the band selection to improve the classification accuracy rate.

We also propose to use simulated annealing with geometric cooling schema. A new objective function is designed based on two important terms, the first is a score measured under 100 decision trees (bagging) and the second is the correlation between the bands.

The performance of the proposed approach was tested under three different Hyperspectral images: Salinas, Pavia University and Indian Pines.

The analysis of the results obtained by the proposed approach demonstrates the outperformance of the proposed approach.

The proposed approach has provided very good results when it was used in hyperspectral images. In future works, we will propose to improve the approach and the quality of the objective function to adapt it for a different types of datasets.

References

1. **Abdoos, A. A., Khorshidian-Mianaei, P., Rayatpanah-Ghadikolaei, M. (2016).** Combined VMD-SVM based feature selection method for classification of power quality events. Applied Soft Computing, Vol. 38, pp. 637–646. DOI: 10.1016/j.asoc.2015.10.038.

2. **Apolloni, J., Leguizamon, G., Alba, E. (2016).** Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. Applied Soft Computing, Vol. 38, pp. 922–932. DOI: 10.1016/j.asoc.2015.10.037.

3. **Chang, C. I., Liu, K. H. (2014).** Progressive band selection of spectral unmixing for hyperspectral imagery. Geoscience and Remote Sensing, IEEE Transactions on, Vol. 52, No. 4, pp. 2002–2017. DOI: 10.1109/tgrs.2013.2257604.

4. **Garro, B. A., Rodríguez, K., Vázquez, R. A. (2016).** Classification of DNA microarrays using artificial neural networks and ABC algorithm. Applied Soft Computing, Vol. 38, pp. 548–560. DOI: 10.1016/j.asoc.2015.10.002.

5. **Gold, C., Holub, A., Sollich, P. (2005).** Bayesian approach to feature selection and parameter tuning for support vector machine classifiers. Neural Networks, Vol. 18, No. 5–6, pp. 693–701.

6. **Kohavi, R., John, G. H. (1997).** Wrappers for feature subset selection. Artificial Intelligence, Vol. 97, No. 1–2, pp. 273–324.

7. **Lin, Y., Hu, Q., Liu, J., Chen, J., Duan, J. (2016).** Multi-label feature selection based on neighborhood mutual information. Applied Soft Computing, Vol. 38, pp. 244–256.

8. **Medjahed, S. A., Saadi, T. A., Benyettou, A. (2016).** Microcanonical annealing and threshold accepting for parameter determination and feature selection of support vector machines. Journal of Computing and Information Technology, Vol. 24, No. 4, pp. 369–382. DOI: 10.20532/cit.2016.1003342.

9. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2016).** Gray wolf optimizer for hyperspectral band selection. Applied Soft Computing, Vol. 40, pp. 178–186. DOI: 10.1016/j.asoc.2015.09.045.

10. **Medjahed, S. A., Saadi, T. A., Benyettou, A., Ouali, M. (2017).** Kernel-based learning and feature selection analysis for cancer diagnosis. Applied Soft Computing, Vol. 51, pp. 39–48. DOI: 10.1016/j.asoc.2016.12.010.

11. **Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953).** Equation of state calculations by fast computing machines. Journal of Chemical Physics, Vol. 21, No. 6, pp. 1087–1092. DOI: 10.1063/1.1699114.

12. **Perez-Rodriguez, J., Arroyo-Pena, A. G., Garcia-Pedrajas, N. (2015).** Simultaneous instance and feature selection and weighting using evolutionary computation: Proposal and study. Applied Soft Computing, Vol. 37, pp. 416–443. DOI: 10.1016/j.asoc.2015.07.046.

13. **Sheikhpoura, R., Sarrama, M. A., Sheikhpourb, R. (2016).** Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. Applied Soft Computing, Vol. 40, pp. 113–131. DOI: 10.1016/j.asoc.2015.10.005.

14. **Sun, K., Geng, X., Ji, L., Lu, Y. (2014).** A new band selection method for hyperspectral image based on data quality. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 7, No. 6, pp. 2697–2703. DOI: 10.1109/JSTARS.2014.2320299.

15. **Yuan, Y., Zhu, G., Wang, Q. (2015).** Hyperspectral band selection by multitask sparsity pursuit. Geoscience and Remote Sensing, IEEE Transactions on, Vol. 53, No. 2, pp. 631–644. DOI: 10.1109/TGRS.2014.2326655.