

DIFFERENT OPTIMAL BAND SELECTION OF HYPERSPECTRAL IMAGES USING A CONTINUOUS GENETIC ALGORITHM

S. Talebi Nahr^{a,*}, P. Pahlavani^b, M. Hasanlou^b

^a Department of Civil and Geomatics Engineering, Tafresh University, Postal Code 39518-79611, Tafresh, Iran

^b Center of Excellence in Geomatics Eng. in Disaster Management, Dept. of Surveying and Geomatics Eng., College of Eng., University of Tehran, Tehran, Iran

KEY WORDS: Classification, Band Selection, Hyperspectral Image, Continuous Genetic Algorithm, SVM

ABSTRACT:

In the most applications in remote sensing, there is no need to use all of available data, such as using all of bands in hyperspectral images. In this paper, a new band selection method was proposed to deal with the large number of hyperspectral images bands. We proposed a Continuous Genetic Algorithm (CGA) to achieve the best subset of hyperspectral images bands, without decreasing Overall Accuracy (OA) index in classification. In the proposed CGA, a multi-class SVM was used as a classifier. Comparing results achieved by the CGA with those achieved by the Binary GA (BGA) shows better performances in the proposed CGA method. At the end, 56 bands were selected as the best bands for classification with OA of 78.5 %.

1. INTRODUCTION

In recent years, there are many remotely sensed data available in Photogrammetry and Remote Sensing study fields (Zhang et al., 2009). Hyperspectral remotely sensed (HRS) imagery, LiDAR data, high resolution images, and Radar data are examples of these data. In almost applications, there is no need to use all of these data. However, extracting suitable and adequate features from these data are vital for any remote sensing analysis and especially for the classifying techniques. To overcome these issues, dimension reduction methods are proposed and have shown a direct effect on improving the classifier performance. In this regard, “feature extraction” or “feature selection” are two important terms used in remote sensing fields. In this case, optimal band selection of HRS imagery seems to be crucial to decrease expense of data processing for some tasks such as object detection, pattern recognition, classification, anomaly detection, and etc. Land cover classification is one of the several tasks in remote sensing, which is concerned with the identification of all the different coverages of the earth surface (Chang, 2007; Plaza et al., 2009). HRS imagery has information about materials on earth surface which allows materials to be identified with more accuracy (Chang, 2007; Plaza et al., 2009). Classification of HRS imagery is a challenging task due to its small set of referenced data and the large amount of spectral information (Benediktsson et al., 2005; Tarabalka, 2010), as well as the density estimation (Plaza et al., 2009).

In order to dealing with above difficulties, some researchers apply feature selection/extraction/representation methods (Bazi and Melgani, 2006; Kuo et al., 2009; Serpico and Bruzzone, 2001).

There are some feature extraction methods such as principle component analysis (PCA) or K-L transform that are usually used in the dimension reduction (Green et al., 1988). Fused feature extraction usually makes the interpretation of extracted features more sophisticated. This is why sometimes a fused feature extraction method is not suitable, and a feature selection

method should be used for choosing an optimal subset of original features (bands) (Conese and Maselli, 1993). In most cases, there is some low-frequency information in the bands (noisy bands) that consequently reduce the accuracy of classification of some small classes (Zhang and Pazner, 2007). In addition, when the bands are not correlative with each other, algorithms such as PCA or K-L may not be efficient in the reduction of data dimension. In conclusion, a robust feature/band selection method is important in the classification or preprocessing of HRS imagery (Zhang et al., 2009). In this paper, it is tried to deal with feature selection as a global optimization problem.

Natural evolutionary theory is the fundamental concept of Genetic Algorithms (GAs) (Pal et al., 1998; Ribeiro Filho et al., 1994). GAs are suitable for solving complex optimization problems and for applications that require adaptive problem solving strategies (Brookes, 2001). The effectiveness of GAs search for ideal solutions make GAs probable to be widely utilized in pattern recognition analyses and multi-criteria decisions since 1990s. Also, GAs have been incorporated in the design of feature selection methods (Kudo and Sklansky, 2000).

The Continuous GA (CGA), which is used in this paper, is inherently faster than the binary GA, because the chromosomes do not have to be decoded prior to the evaluation of the cost function.

To reach HRS optimal bands for classification purpose, in this paper, overall accuracy was regarded as a fitness function of the proposed CGA. In this regard, to calculate this function, a classifier would be needed. A Multiclass Support Vector Machine (Multiclass-SVM) was used as the classifier.

This study is organized as follows. The Multiclass-SVM as a classifier to calculate the CGA fitness function is explained in Section 2. In Section 3, the proposed CGA would be illustrated. Study area is introduced in Section 4. The experimental results

* Email: eng.siamaktalebi@gmail.com
Mobile Phone: +98 - 9149718713

are presented in Section 5, and in Section 6, the conclusion is represented.

2. MULTICLASS-SVM AS A CLASSIFIER FUNCTION

The Support Vector Machines (SVM) methodology is based on class separation through margins in which samples are mapped to a feature space where they can be linearly separable (Duda et al., 1995). The data is transformed to a new feature space, generally larger than the original, by a kernel function. Some popular kernels are: Linear, Polynomial and Radial Basis Function (RBF). The ability of separating data with nonlinear distributions is related to the choice of this function and should be chosen according to the problem domain (Duda et al., 1995). Using an appropriate nonlinear mapping, samples of two classes can then be linearly separated by a hyperplane (Bishop, 2006; Duda et al., 1995) in this new transformed and high feature space (Camps-Valls and Bruzzone, 2005; Duda et al., 1995). Thus, SVM training consists of finding an optimal hyperplane where the separating distance between margins of each class can be maximized (Bishop, 2006; Duda et al., 1995). Samples whose locations are located on the margins are called support vectors and are the most informative ones to build the classification decision boundary (Duda et al., 1995).

Multiclass-SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The dominant approach for doing so is to reduce the single multiclass problem into multiple binary classification problems (Duan and Keerthi, 2005). Common methods for such reduction include (Duan and Keerthi, 2005; Hsu and Lin, 2002):

1. Building binary classifiers which distinguish between (i) one of the labels and the rest (one-versus-all) or (ii) between every pair of classes (one-versus-one). Classification of new instances for the one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class (it is important that the output functions be calibrated to produce comparable scores). The one-versus-all approach builds n SVM models where n is the number of classes. The i^{th} SVM is trained with all of the examples in the i^{th} class with positive labels, and all other examples with negative labels. Thus, given l training data $(x_1, y_1), \dots, (x_l, y_l)$, where $x_i \in R^n, i = 1, \dots, l$ and $y_i \in \{1, \dots, \kappa\}$ is the class of x_i , the i^{th} SVM solves the following problem:

$$\begin{aligned} \min_{\omega^i, b^i, \xi^i} \quad & \frac{1}{2} (\omega^i)^T \omega^i + C \sum_{j=1}^l \xi_j^i (\omega^i)^T \\ & (\omega^i)^T \Phi(x_j) + b^i \geq 1 - \xi_j^i, \quad \text{if } y_j = i \\ & (\omega^i)^T \Phi(x_j) + b^i \leq 1 + \xi_j^i, \quad \text{if } y_j \neq i \\ & \xi_j^i \geq 0, \quad j = 1, \dots, l \end{aligned} \quad (1)$$

where the training data x_i are mapped to a higher dimensional space by the function Φ , and C is the penalty parameter.

For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification. This method

constructs $n(n-1)/2$ classifiers where each one is trained on data from two classes. For training data from the i^{th} and the j^{th} classes, we solve the following binary classification problem:

$$\begin{aligned} \min_{\omega^{ij}, b^{ij}, \xi^{ij}} \quad & \frac{1}{2} (\omega^{ij})^T \omega^{ij} + C \sum_t \xi_t^{ij} (\omega^{ij})^T \\ & (\omega^{ij})^T \Phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \quad \text{if } y_t = i \\ & (\omega^{ij})^T \Phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \quad \text{if } y_t = j \\ & \xi_t^{ij} \geq 0. \end{aligned} \quad (2)$$

2. Directed acyclic graph SVM (DAGSVM) (Platt et al., 1999).
3. Error-correcting output codes (Dietterich and Bakiri, 1995).

Crammer and Singer proposed a multiclass-SVM method which casts the multiclass classification problem into a single optimization problem, rather than decomposing it into multiple binary classification problems (Crammer and Singer, 2002). See also Lee, Lin and Wahba (Lee et al., 2001, 2004).

3. THE PROPOSED CONTINUOUS GENETIC ALGORITHM

Genetic Algorithms (GAs) are the most widespread techniques among evolutionary algorithms. They allow us to find potential solutions for optimization problems in an acceptable time, especially when the search space is very large (Goldberg and Holland, 1988). This technique is heuristic, based on population of individuals (e.g., chromosomes), in which each individual enacts a candidate solution for a problem (Goldberg and Holland, 1988) and can be represented as a bit string. Each individual is evaluated by a function called fitness. This function establishes the quality of an individual related to a solution.

In CGA, the population starts in a random way, or by some strategy according to the problem in question. The population undergoes a determined number of evolutions. During this process individuals of this population are evolved and reproduced using some genetic operations such as: crossover, mutation and the selection process. Its main goal is to find the individual with the best fitness (Goldberg and Holland, 1988).

We modeled the problem of band selection as follows: each individual has a size of B bands/genes, as shown in Figure 1, and each gene represents a uniformly random double vector number between -1 and 1, associated with a band. As previously stated, the quality of each candidate solution is evaluated according to a fitness function. Our fitness function here is the Overall Accuracy (OA) of a classification performed by a multiclass-SVM. Stochastic Universal Sampling was used as the selection operator. Moreover, Uniform crossover, as well as Gaussian mutation were used as CGA operators. Finally, migration direction was set to forward mode in addition to the elitism mechanism.

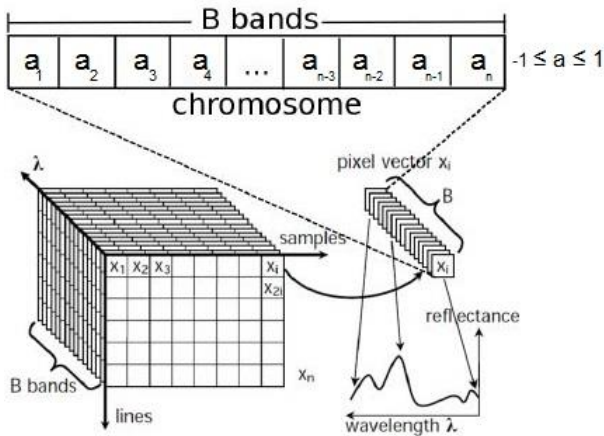


Figure 1: Chromosome representation, adapted from (Tarabalka, 2010).

4. STUDY AREA

One of the most widely used hyperspectral images in the literature (Fu et al., 2006; Jia et al., 2010; Martínez-Usó et al., 2007; Qian et al., 2011; Sarhrouni et al., 2012) was used in this study (Figure 2). This data set was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range of 0.4–2.5 μm .



Figure 2: Study area image in false colours.

5. EXPERIMENTAL RESULTS

In this section, we describe the experiments carried out on a well-known Indian Pines dataset for validating our proposed method in finding the most suitable features. In this paper, the C++ external library was used to program an executive-file for implementing multiclass-SVM on a PC with a 8 GB of RAM, and a 2.20-GHz-based processor. Details regarding to CGA and classifiers setups would be presented in coming section.

5.1 CGA and Multiclass-SVM setups

An individual's chromosome, i.e., the features present in an individual, was initialized in a random way and the parameters were set according to results of preliminary experiments. Table 1 presents all parameters set in CGA. The samples were randomly chosen, however the total number of samples has an important impact on the performance. The higher the number of samples are, the higher the time to calculate the fitness for each individual would be. In order to ensure high reliability of results, 10 runs of CGA for each dataset were performed. Then those features that appeared more frequently were selected. The kernel used in Multiclass-SVM was the Radial Basis Function (RBF).

Parameters	Value
Population size	20
Number of Generations	470
Crossover rate	0.8
Migration Fraction	0.2
Migration Interval	20
Scale (Gaussian Mutation)	0.5
Shrink (Gaussian Mutation)	0.7
Elite count	1

Table 1: GA's Parameters

5.2 Selected features

Since we have decided to let CGA find the optimal number of features, it is also important to note that for each run we obtained almost the same number of features. This indicates that there are tracks of spectral bands that are more discriminative than others. Thereby, features that repeat in at least 50% of all 10 runs were selected to compose the final subset of features. The number of selected features for our proposed CGA was 56 and for Binary GA (BGA) was 86 (Santos et al., 2012), which are selected automatically. Figure 3 shows the best and the mean of fitness values in each generation for a single run of CGA.

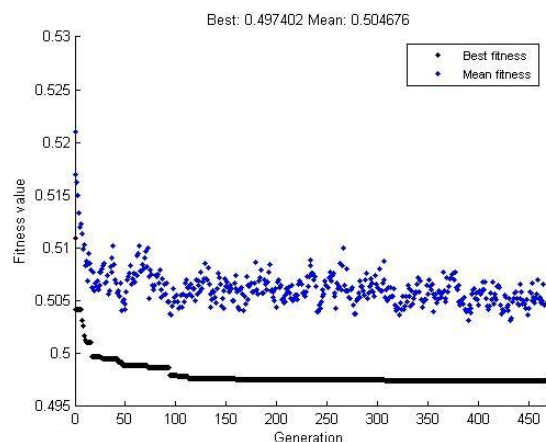


Figure 3: Fitness values achieved by the proposed CGA.

5.3 Evaluation

As shown in Table 2, the results of our proposed CGA method were close to the widely used BGA, but, improvements in each of classes are noticeable. Results for BGA are derived from (Santos et al., 2012). Because of reduced feature space in CGA approach, classification process is less computationally expensive than BGA. Figure 4 shows the classification map.



Figure 4: (a) The achieved classification map, (b) Ground truths.

6. CONCLUSION AND FUTURE WORK

Due to the high dimensionality of remote sensed hyperspectral images, their rich content and some drawbacks for data discrimination, we have investigated the benefits of using feature selection approaches for the problem of classification of this data type. Thus, in this paper, a new filter feature selection approach was proposed. The main idea in this proposal was that smaller subsets of features, which generate high values of OA, may provide enough discriminant information for the classification task and hence it would be easier to build decision boundaries with good power of generalization. In this regard, we used the advantages of continuous genetic algorithms to lead a search for finding better clusters from a subset of features

	Number of Samples	SVM-GA (%)	SVM-CGA (%)
Alfalfa	46	90.7	95.5
Corn-notill	1428	63.3	85.6
Corn-mintill	830	53.2	89.2
Corn	237	23.5	83.1
Grass-pasture	483	83.1	98.8
Grass-trees	730	95.2	95.4
Grass-pasture-mowed	28	88.5	0.0
Hay-windrowed	478	93.5	95.2
Oats	20	100	0.0
Soybean-notill	972	60.5	79.2
Soybean-mintill	2455	84.9	79.8
Soybean-clean	593	83.0	87.0
Wheat	205	93.4	95.5
Woods	1265	98.5	94.4
Buildings-Grass-Trees-Drives	386	39.5	100
Stone-Steel-Towers	93	98.9	100
OA		77.1	78.5

Table 2: Comparison of a classification performed using CGA and BGA (Santos et al., 2012).

REFERENCES

Bazi, Y., Melgani, F., 2006. Toward an optimal SVM classification system for hyperspectral remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on* 44, 3374-3385.

Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. *Geoscience and Remote Sensing, IEEE Transactions on* 43, 480-491.

Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer New York.

Brookes, C.J., 2001. A genetic algorithm for designing optimal patch configurations in GIS. *International Journal of Geographical Information Science* 15, 539-559.

Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on* 43, 1351-1362.

Chang, C.I., 2007. *Hyperspectral data exploitation: theory and applications*. Wiley-Interscience.

Conese, C., Maselli, F., 1993. Selection of optimum bands from TM scenes through mutual information analysis. *ISPRS journal of photogrammetry and remote sensing* 48, 2-11.

Crammer, K., Singer, Y., 2002. On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research* 2, 265-292.

Dietterich, T.G., Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. *arXiv preprint cs/9501101*.

Duan, K.-B., Keerthi, S.S., 2005. Which is the best multiclass SVM method? An empirical study, *Multiple Classifier Systems*. Springer, pp. 278-285.

Duda, R.O., Hart, P.E., Stork, D.G., 1995. *Pattern Classification and Scene Analysis* 2nd ed. Wiley Interscience.

Fu, Z., Caelli, T., Liu, N., Robles-Kelly, A., 2006. Boosted band ratio feature selection for hyperspectral image classification, *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. IEEE, pp. 1059-1062.

Goldberg, D.E., Holland, J.H., 1988. *Genetic algorithms and machine learning*. Machine learning 3, 95-99.

Green, A.A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions on* 26, 65-74.

- Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on* 13, 415-425.
- Jia, S., Qian, Y., Li, J., Liu, W., Ji, Z., 2010. Feature extraction and selection hybrid algorithm for hyperspectral imagery classification, *Geoscience and Remote Sensing Symposium (IGARSS)*, 2010 IEEE International. IEEE, pp. 72-75.
- Kudo, M., Sklansky, J., 2000. Comparison of algorithms that select features for pattern classifiers. *Pattern recognition* 33, 25-41.
- Kuo, B.-C., Li, C.-H., Yang, J.-M., 2009. Kernel nonparametric weighted feature extraction for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on* 47, 1139-1155.
- Lee, Y., Lin, Y., Wahba, G., 2001. Multicategory support vector machines, *Proceedings of the 33rd Symposium on the Interface*. Citeseer.
- Lee, Y., Lin, Y., Wahba, G., 2004. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association* 99, 67-81.
- Martínez-Usó, A., Pla, F., Sotoca, J.M., García-Sevilla, P., 2007. Clustering-based hyperspectral band selection using information measures. *Geoscience and Remote Sensing, IEEE Transactions on* 45, 4158-4171.
- Pal, S.K., Bandyopadhyay, S., Murthy, C., 1998. Genetic algorithms for generation of class boundaries. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 28, 816-828.
- Platt, J.C., Cristianini, N., Shawe-Taylor, J., 1999. Large Margin DAGs for Multiclass Classification, *nips*, pp. 547-553.
- Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment* 113, S110-S122.
- Qian, Y., Zhou, J., Ye, M., Wang, Q., 2011. Structured sparse model based feature selection and classification for hyperspectral imagery, *Geoscience and Remote Sensing Symposium (IGARSS)*, 2011 IEEE International. IEEE, pp. 1771-1774.
- Ribeiro Filho, J.L., Treleaven, P.C., Alippi, C., 1994. Genetic-algorithm programming environments. *Computer* 27, 28-43.
- Santos, A., Celes, C.d.S., Araújo, A.d.A., Menotti, D., 2012. Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithm and cluster validity, *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV'12)*, pp. 675-681.
- Sarhrouni, E., Hammouch, A., Aboutajdine, D., 2012. Band Selection and Classification of Hyperspectral Images using Mutual Information: An algorithm based on minimizing the error probability using the inequality of Fano, *Multimedia Computing and Systems (ICMCS)*, 2012 International Conference on. IEEE, pp. 155-159.
- Serpico, S.B., Bruzzone, L., 2001. A new search algorithm for feature selection in hyperspectral remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on* 39, 1360-1367.
- Tarabalka, Y., 2010. Classification of hyperspectral data using spectral-spatial approaches. Ph. D. dissertation, University of Iceland and Grenoble Institute of Technology.
- Zhang, X., Pazner, M., 2007. Comparison of lithologic mapping with ASTER, hyperion, and ETM data in the southeastern Chocolate Mountains, USA. *Photogrammetric engineering and remote sensing* 73, 555.
- Zhang, X., Sun, Q., Li, J., 2009. Optimal band selection for high dimensional remote sensing data using genetic algorithm, *Proc. of SPIE Vol*, pp. 74711R-74711.