J Indian Soc Remote Sens (June 2014) 42(2):423–428
DOI 10.1007/s12524-013-0327-7

SHORT NOTE

ISRS

# Multiple Classifier Systems for Hyperspectral Remote Sensing Data Classification

**Iman Khosravi · Majid Mohammad-Beigi**

Received: 15 April 2013 / Accepted: 12 September 2013 / Published online: 11 December 2013


**Abstract** One of the most widely used outputs of remote sensing technology is Hyperspectral image. This large amount of information can increase classification accuracy. But at the same time, conventional classification techniques are facing the problem of statistical estimation in high-dimensional space. Recently in remote sensing, support vector machines (SVMs) have shown very suitable performance in classifying high dimensionality problem. Another strategy that has recently been used in remote sensing is multiple classifier system (MCS). It can also improve classification accuracy by combining different classifier methods or by a diversity of the same classifier. This paper aims to classify a Hyperspectral data using the most common methods of multiple classifier systems i.e. adaboost and bagging and a MCS based on SVM. The data used in the paper is an AVIRIS data with 224 spectral bands. The final results show the high capability of SVMs and MCSs in classifying high dimensionality data.

**Keywords** Multiple classifier system · Support vector machine · Hyperspectral data classification · Correlation-based feature selection

I. Khosravi (✉)
Department of Surveying Engineering, Faculty of Engineering, University of Isfahan, Isfahan, I.R., Iran
e-mail: iman.khosravi@mihanmail.ir

I. Khosravi
e-mail: civiliman@gmail.com

M. Mohammad-Beigi
Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, I.R., Iran
e-mail: majid.beigi@eng.ui.ac.ir

## Introduction

### Overview

One of the widely utilized outputs of remote sensing technology is the Hyperspectral image. This data covers a wide spectral range from visible to short-wavelength infrared at a large number of spectral channels. In a classification task, whatever the dimensionality of the data increases, the capability of detecting different classes is increased (Fauvel et al. 2008). However, the high number of features and inadequate number of training samples can decrease the accuracy of classification. This problem is known as Hughes phenomenon (Fauvel et al. 2006). In this case, conventional statistical algorithms such as maximum likelihood cannot produce appropriate results and it is necessary to consider other methods (Fauvel et al. 2006).

Up to now, several different classification algorithms have been proposed for solving this problem. Recently in remote sensing, it has been shown that support vector machines (SVMs) has very suitable performance in classifying high dimensional data (e.g. (Camps-Valls and Bruzzone 2005; Ceamanos et al. 2010; Fauvel et al. 2006, 2008; Pal and Mather 2004). The first use of SVMs for classifying remotely sensed images had acceptable results (Gualtieri and Chettri 2000). Pal and Mather (2004) demonstrated that SVMs have high capability of generalization and they do not have the impact of Hughes phenomenon (Pal and Mather 2004). Camps-Valls and Bruzzone (2005) showed SVMs produce equal or better accuracy than other classifiers in high dimensional data classification (Camps-Valls and Bruzzone 2005). Fauvel et al. (2006) considered two different kernels and compared them together. He evaluated the capability of generalization of SVM and showed that SVMs have high ability to deal with high dimensional features space and small training samples (Fauvel et al. 2006).

In addition to SVMs, multiple classifier systems (MCSs) have been also recently used in remote sensing (Benediktsson et al. 2007). For example, Ham et al. (2005) used MCS for hyperspectral images classification and achieved successful results (Ham et al. 2005). Both of the above concepts (SVM & MCS) can also be used together. In (Waske and Benediktsson 2007), SVMs were used for classifying multi-sensor imagery so that each single SVM classifier was trained separately by each source and finally, a decision fusion based on SVMs was used. In (Malleswara Rao et al. 2012), a new hybrid SVM method was developed so that it increased the classification accuracy while lowering the computational cost and complexity of the process. In (Bigdeli et al. 2013), Naive Bayes as a classifier fusion algorithm combined decision of SVM classifiers for hyperspectral data classification.

This study aims to classify a Hyperspectral data using the most common methods of multiple classifier systems and a SVM-based classifier ensemble. First, two best feature subsets are selected from total features. Then, the selected feature subsets were classified to three classes using two multi-class SVM methods. Finally, another classification algorithm is implemented using a multiple classifier system based on multi-class SVM and then by adaboost and bagging.

## Support Vector Machines

Support vector machines method is based on statistical learning theory (Haykin 2005) and recently has been frequently used for remote sensing data classification problems. For a two-class problem in the feature space, SVM procedure consists of finding the hyperplane that maximizes the margin i.e. the distance to the closest training data points in both classes (Fig. 1) (Haykin 2005).

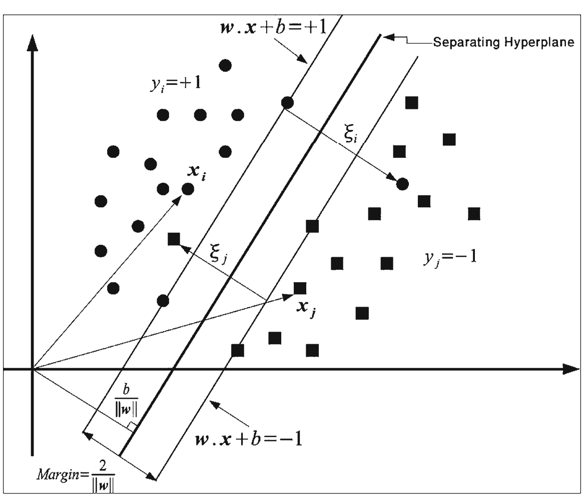


**Fig. 1** Classification by SVMs. Margin should be maximized. Optimum separating hyperplane (Fauvel et al. 2006)

Using the so-called Kernel Trick, one can generalize SVMs to non-linear decision functions. It is necessary to be fulfilled Mercers' conditions by the kernel function (Haykin 2005). The most common used kernels in remote sensing are the polynomial and Gaussian radial basis functions (RBF). They are defined as follow (Haykin 2005):

$$k_{poly}(x_i, x_j) = [(x_i, x_j) + 1]^p \tag{1}$$

$$k_{gauss}(x_i, x_j) = \exp\left[-\gamma \|x_i - x_j\|^2\right] \tag{2}$$

Degree parameter (d) for the polynomial kernel and sigma parameter for the RBF kernel must be specified. Moreover, another parameter (i.e. C) is used to control overfitting (Haykin 2005).

## Multiclass SVM

In some applications of remote sensing such as land-cover/land-use classification, we have more than two classes. Therefore binary classification is less applicable in here. Up to now, several algorithms have been proposed to generate multiclass SVMs from binary SVMs. Two most common algorithms are one-against-one (OAO) and one-against-all (OAA) (Hsu and Lin 2002). In the first algorithm (OAO), SVM classifier is produced for all possible pairs of classes. Thus, M (M – 1)/2 binary SVMs are generated for M classes. For this 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,

**Fig. 2** San Diego urban area

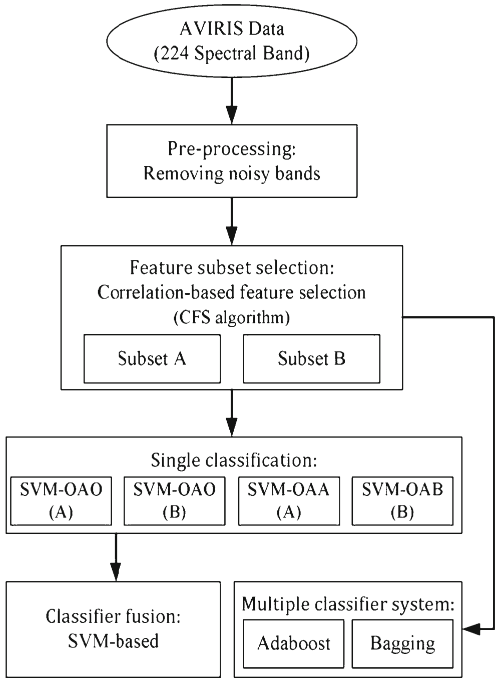


**Fig. 3** Schematic diagram of the methodology

and finally the class with the most votes determines the instance classification. In the second algorithm (OAA), one class is compared with the other classes. Therefore, M binary SVMs are needed for classifying M classes. Classification of test samples for this approach 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) (http://en.wikipedia.org/wiki/Support_vector_machine/).

### Multiple Classifier Systems

Using several different decisions in the final classification, instead of relying on a single classifier is the basic idea of a multiple classifier system (Benediktsson and Kanellopoulos 1999). Other different names for this concept can be found in machine learning and pattern recognition are *committees of learners*, *mixtures of experts*, *classifier ensembles*, *multiple classifier systems*, *consensus theory* and etc. MCSs are based on the supposition that a set of autonomous or diverse classifiers cause individual errors, which are not produced by the majority of the other classifiers. MCSs can be generated by two procedures: by combining variant classifier algorithms (e.g. Ham et al. 2005 and Waske and Vander Linden 2008), or by a composition of variants of a learner, the so-called base classifier (e.g. Waske and Benediktsson 2007). Boosting and bagging are the most common used concepts to produce such a diverse set of classifiers (Haykin 2005).

**Table 1** Two feature subsets selection by CFS algorithm

| Subset | Features | Number |
|---|---|---|
| A | Bands{1, 2, 4, 7, 8, 9, 11, 17, 23, 24, 25, 42, 53, 61, 78, 113, 123, 137, 143, 161, 182, 185} | 22 |
| B | Bands{15, 20, 22, 28, 30, 33, 48, 54, 62, 76, 98, 101, 104, 116, 136, 147, 151, 186, 189} | 19 |

**Table 2** The optimum values of RBF kernel parameters

| Kernel Parameters | Subset A | Subset B |
|---|---|---|
| C (for tuning) | 10,000 | 1,000 |
| Gama (for Gaussian) | 0.0003 | 0.007 |

## Study Area and Dataset

The data used in this paper is a Hyperspectral AVIRIS data acquired over an air station at San Diego, California. The image is available as an example data in *ENVI* software (Fig. 2). AVIRIS sensor has 224 spectral channels from 400 to 2,500 nm. Its spatial resolution depends on height of ER-2 aircraft and usually changes from 4 to 20 m. In this study, three classes are considered: ground, vegetation and building regions. Number of samples for each class is as follow: ground: 2,063, vegetation: 1,668 and building: 2,256 pixels.

## Methodology

The proposed method uses from both SVM and MCS concepts for classifying Hyperspectral data. The schematic diagram of the proposed method is shown in Fig. 3. According to the figure, it has these steps: first of all, AVIRIS data is preprocessed. Some spectral bands that are noisy should be removed. Then, two feature subsets are selected using correlation-based feature selection (CFS) algorithm. The CFS technique assumes that useful feature subsets contain features that are predictive of the class but uncorrelated with one another. CFS computes a heuristic measure of the “merit” of a feature subset from pair-wise feature correlations and a formula adapted from test theory. Heuristic search is used to traverse the space of feature subsets in reasonable time; the subset with the highest merit found during the search is reported (Hall 1998).

Next, both multi-class SVM methods i.e. SVM-OAO and SVM-OAA are separately applied to each feature subset.

**Table 3** The overall accuracy of four single SVM classifiers

| Overall accuracy | SVM-OAO | SVM-OAA |
|---|---|---|
| Subset A | 97.23 % | 89.57 % |
| Subset B | 96.81 % | 91.04 % |

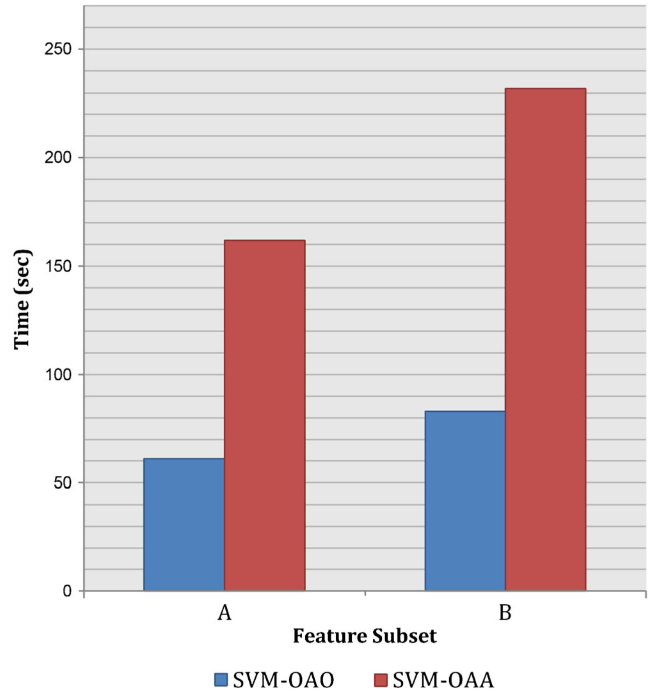


**Fig. 4** The comparison of four single SVM classifiers in terms of time (*black bar*: OAO and *gray bar*: OAA)

Finally, the rule images resulting from every SVM are fused (i.e. combined and classified) by an additional SVM in accordance to (Waske and Benediktsson 2007). The same samples are used throughout the classification process. In contrast to (Waske and Benediktsson 2007), a single Hyperspectral source is now available instead of the multi-source data. In addition, other classifier ensembles i.e. adaboost and bagging (with decision tree as base classifier) are used. Finally, the results of all these methods are compared in terms of overall accuracy and time.

## Implementation and Results

First, 35 bands of data must be removed from data due to water absorption and bad bands. Thus, 189 bands are remained. Then, two best feature subsets from total features are selected using CFS algorithm (with scatter search (Lopez et al. 2006)). Table 1 shows the feature subsets (selected bands) by this algorithm.

**Table 4** The overall accuracy and time of SVM classifiers for PCA and ICA features

| Overall accuracy | SVM-OAO | SVM-OAA |
|---|---|---|
| PCA | 92.31 % | 85.86 % |
| | 25 s | 110 s |
| ICA | 97.34 % | 88.13 % |
| | 105 s | 315 s |

After selecting feature subsets, both multi-class SVM methods are applied to each feature subset. Therefore, four classifiers namely SVM-OAO-(A), SVM-OAO-(B), SVM-OAA-(A) and SVM-OAA-(B) will be generated. In this study, only RBF kernel is used for all SVMs. In addition, the optimum values for tuning and Gaussian parameters (C and gamma parameters) for each feature subset are determined beforehand. These values are shown in Table 2.

For evaluating the SVMs, K-fold cross validation method is used, here. In K-fold cross validation, total classified samples are divided into K equal parts so that one part is as test sample and K – 1 other part is considered as training samples. Here, the data are divided into $k=5$ parts. The overall accuracy of classifications achieved by four SVMs is shown in Table 3.

In addition, Fig. 4 shows the performance of each SVM classifier in terms of time. According to Table 3, SVM-OAO classifier could produce 6 % to 7 % higher overall accuracy (97.23 % and 96.81 %) than that of SVM-OAA classifier (89.57 % and 91.04 %) for both subsets. Furthermore,

**Table 5** The overall accuracy and time of MCS methods

| | MCS based on SVM | AdaBoost (J48) | Bagging (J48) |
|---|---|---|---|
| $K=5$ | 99.82 % | 99.82 % | 99.78 % |
| | 9 s | 80 s | 59 s |
| $K=10$ | 99.82 % | 99.85 % | 99.80 % |
| | 16 s | 130 s | 123 s |
| $K=20$ | 99.85 % | 99.85 % | 99.82 % |
| | 28 s | 265 s | 244 s |

SVM-OAO needs less time than SVM-OAA for classifying image. However, the number of classifiers created by SVM-OAA is generally much larger than the SVM-OAA algorithm. Although, the number of samples data needed for each classifier is much smaller.

In order to realize the performance of CFS algorithm, two common feature extraction algorithms i.e. principle component analysis (PCA) and independent component analysis (ICA) are used in this paper. PCA produces new features from initial features so that their variances are maximum, whereas ICA maximizes joint entropy and minimizes mutual information between features (Haykin 2005). After extracting new features by PCA and ICA, both multi-class SVM methods are applied to each of them. Table 4 shows the overall accuracy and needed time for four SVMs.

From Tables 3 and 4, it can be concluded that the overall accuracy of SVMs with CFS algorithm are better when compared to two SVMs with PCA (around 5 % at both OAO and OAA modes). Conversely, the SVMs with PCA need less time than the SVMs with CFS. In addition, SVM-OAO with ICA could produce higher overall accuracy than SVM-OAO with CFS. However, it needs more time than SVM-OAO with CFS for classifying image. Generally speaking, the overall accuracy values of four SVMs with CFS are better than the other two algorithms. In addition, the time needed for SVMs with CFS is much less than that of SVMs with ICA, but it is much higher than SVMs with PCA.

In next step, the rule images resulting from every SVM are fused by an additional SVM. In addition, other classifier ensembles i.e. *adaboost* and *bagging* (with decision tree as base classifier) are used. For evaluating performance of the methods, K-fold cross validation is used with K=5, 10 and 20. Table 5 shows the overall accuracy and needed time for each method.

According to Tables 3 and 5, multiple classifier system based on SVM could produce between 2 % to 10 % higher overall accuracy than single SVM classifiers. In addition, MCS based on SVM needs less time than single SVM classifiers. Other MCS methods, i.e. adaboost and bagging could also produce higher overall accuracy than single SVM classifiers. Among these algorithms, MCS based on SVM needs minimum time for classifying image.

## Conclusion

This paper focused on the Hyperspectral image classification using both SVM and MCS concepts. Two best feature subsets were selected and then, two multi-class SVM classifiers i.e. SVM-OAO and SVM-OAA are applied. The results demonstrated that SVM-OAO classifier could produce higher overall accuracy compared with SVM-OAA classifier at both feature subsets. In addition, SVM-OAO classifier needed less time than SVM-OAA for classification. However, MCS based on SVM could produce higher overall accuracy and it needed less time than single SVM classifiers. Furthermore, adaboost and bagging could produce higher accuracy than single SVM classifiers, although base classifier of each was decision tree. This issue showed the high ability of MCSs that they could produce high accuracy from fusing of weak learners. Among all algorithms, MCS based on SVM needed minimum time for classifying image. Finally, this paper concludes SVM and MCS approaches are not affected by Hughes phenomenon and they can be efficient methods for classifying high-dimensional data when even there are inadequate samples.

## References


Available online at: http://en.wikipedia.org/wiki/Support_vector_machine/. Accessed 30 December 2012.

Benediktsson, J. A., & Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing, 37*(3), 1367–1377.

Benediktsson, J. A., Chanussot, J., & Fauvel, M. (2007). Multiple classifier systems in remote sensing: from basics to recent developments. *MCS'07 Proceedings of the 7th international conference on multiple classifier systems*. Heidelberg, Germany: Springer.

Bigdeli, B., Samadzadegan, F., & Reinartz, P. (2013). A multiple SVM system for classification of Hyperspectral remote sensing data. *Journal of the Indian Society of Remote Sensing, May 2013.* doi: 10.1007/s12524-013-0286-z.

Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 43*(6), 1351–1362.

Ceamanos, X., Waske, B., Benediktsson, J. A., Chanussot, J., Fauvel, M., & Sveinsson, J. R. (2010). A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion, 1*(4), 293–307.

Fauvel, M., Chanussot, J., & Benediktsson, J. A. (2006). A combined support vector machines classification based on decision fusion. *IEEE International on Geoscience and Remote Sensing Symposium*, 2494–2497.

Fauvel, M., Benediktsson, J. A., Chanussot, J., & Sveinsson, J. R. (2008). Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing, 46*(11), 4834–4837.

Gualtieri, J. A., & Chettri, S. (2000). Support vector machines for classification of hyperspectral data. *in Proc. IGARSS, Honolulu, HI, 2*, 813–815.

Hall, M. A. (1998). Correlation-based feature selection for machine learning, *PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand*.

Ham, J., Chen, Y., Crawford, M. M., & Gosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 43*(3), 492–501.

Haykin, S. (2005). *Neural networks: A comprehensive foundation*. Prentice-Hall, 340–360.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 13*(2), 415–425.

Lopez, F. G., Torres, M. G., Batista, B. M., Perez, J. A. M., & Moreno-Vega, J. M. (2006). Solving feature subset selection problem by a parallel scatter search. *European Journal of Operational Research, 169*, 477–489.

Malleswara Rao, T. C., Jai Sankar, G., & Roopesh Kumar, T. (2012). A hierarchical hybrid SVM method for classification of remotely sensed data. *Journal of the Indian Society of Remote Sensing, 40*(2), 191–200.

Pal, M., & Mather, P. M. (2004). Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems, 20*(7), 1215–1225.

Waske, B., & Benediktsson, J. A. (2007). Fusion of support vector machines for classification of multisensor data. *IEEE Transaction on Geoscience and Remote Sensing, 45*(12), 3858–3866.

Waske, B., & Vander Linden, S. (2008). Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transactions on Geoscience and Remote Sensing, 46*(5), 1457–1466.