ISPRS Journal of Photogrammetry and Remote Sensing 66 (2011) 247–259

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Review article

# Support vector machines in remote sensing: A review

Giorgos Mountrakis *, Jungho Im, Caesar Ogole

*Department of Environmental Resources Engineering, SUNY College of Environmental Science and Forestry, 1 Forestry Dr, Syracuse, NY 13210, USA*

ARTICLE INFO

*Article history:*
Received 6 June 2010
Received in revised form
17 September 2010
Accepted 1 November 2010
Available online 3 December 2010

*Keywords:*
Support vector machines
Review
Remote sensing
SVM
SVMs

ABSTRACT

A wide range of methods for analysis of airborne- and satellite-derived imagery continues to be proposed and assessed. In this paper, we review remote sensing implementations of support vector machines (SVMs), a promising machine learning methodology. This review is timely due to the exponentially increasing number of works published in recent years. SVMs are particularly appealing in the remote sensing field due to their ability to generalize well even with limited training samples, a common limitation for remote sensing applications. However, they also suffer from parameter assignment issues that can significantly affect obtained results. A summary of empirical results is provided for various applications of over one hundred published works (as of April, 2010). It is our hope that this survey will provide guidelines for future applications of SVMs and possible areas of algorithm enhancement.



## 1. Introduction

Remotely-sensed data are used in numerous applications. Typically, an image classification process is initiated to convert data into meaningful information. Unfortunately, image classification is not a trivial task. As noted by Chi et al. (2008), classification of remote sensing data is particularly daunting because most of the supervised learning schemes require sufficiently large amount of training samples, yet definition and acquisition of reference data is often a critical problem. Various classification techniques, both parametric and non-parametric, have been developed and used in different contexts — remote sensing inclusive.

Previous reviews, such as that by Plaza et al. (2009), focused on recent developments in methodologies for processing a specific type of imagery, for example hyperspectral images. The review provided in this paper follows the algorithmic perspective rather than image characteristics. More specifically, we focus on applications of support vector machines (SVMs) in remote sensing. The motivation to carry out this study comes from different sources. First, SVMs are not as well-known as other classifiers (e.g., decision trees, variants of neural networks) in the general remote sensing community, yet they can match if not exceed the performance of established methods. Second, their performance gains seem well-suited for remote sensing applications, where a limited amount of reference data is often provided. Third, even though the method is not widely popular, in recent years there has been a significant increase in SVM works on remote sensing problems suggesting this review is current and appropriate.

This review focuses on recent research papers (available by April, 2010) published in eight major journals of remote sensing, namely, ISPRS Journal of Photogrammetry and Remote Sensing, Remote Sensing of Environment, Photogrammetric Engineering & Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, International Journal of Remote Sensing, Canadian Journal of Remote Sensing and GIScience and Remote Sensing. A limited number of research papers relevant to the thematic point and thus included in this review came from additional sources. The selected papers represent a wide range of: (i) applications from coal reserve detection to urban growth monitoring, (ii) resolutions from sub-meter to several kilometers pixel size, (iii) spectral resolution from single to hundreds of bands, and (iv) comparative methods from maximum likelihood classifiers to neural networks. For completeness, we first recap on the basics of SVM methodology before diving into specific works. Relevant papers are then summarized, while juxtaposition of general patterns enables us to derive conclusions and recommendations for further investigations.

## 2. Overview of support vector machines

Support vector machines (SVMs) is a supervised non-parametric statistical learning technique, therefore there is no assumption

* Corresponding address: Department of Environmental Resources Engineering, SUNY College of Environmental Science and Forestry, 419 Baker Hall, 1 Forestry Dr, Syracuse, NY 13210, USA. Tel.: +1 (315) 470 4824; fax: +1 (315) 470 6958.
*E-mail address:* gmountrakis@esf.edu (G. Mountrakis).
*URL:* http://www.aboutgis.com (G. Mountrakis).

0924-2716/$ – see front matter © 2010 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2010.11.001

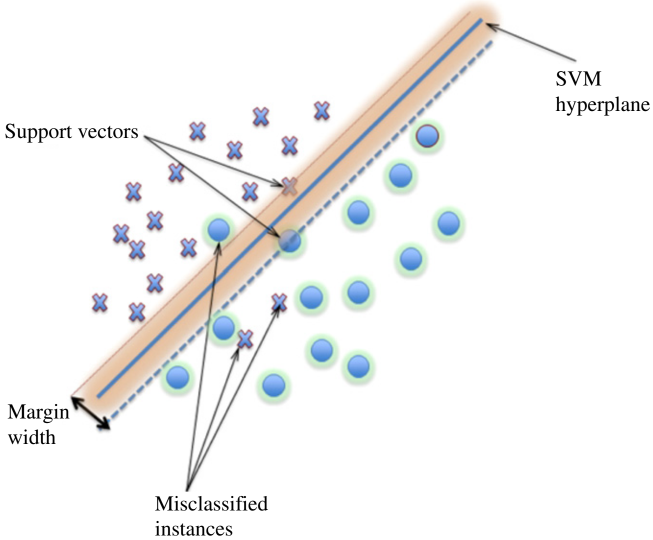


**Fig. 1.** Linear support vector machine example.
*Source:* adapted from Burges (1998).

made on the underlying data distribution. In its original formulation (Vapnik, 1979) the method is presented with a set of labeled data instances and the SVM training algorithm aims to find a hyperplane that separates the dataset into a discrete predefined number of classes in a fashion consistent with the training examples. The term optimal separation hyperplane is used to refer to the decision boundary that minimizes misclassifications, obtained in the training step. Learning refers to the iterative process of finding a classifier with optimal decision boundary to separate the training patterns (in potentially high-dimensional space) and then to separate simulation data under the same configurations (dimensions) (Zhu and Blumberg, 2002).

In its simplest form, SVMs are linear binary classifiers that assign a given test sample a class from one of the two possible labels. An instance of a data sample to be labeled in the case of remote sensing classification is normally the individual pixel derived from the multi-spectral or hyperspectral image. Such a pixel is represented as a pattern vector, and for each image band, it consists of a set of numerical measurements. Elements of the feature vector may also include other discriminative variable measurements based on pixel spatial relationships such as texture. Fig. 1 illustrates a simple scenario of a two-class separable classification problem in a two-dimensional input space. An important generalization aspect of SVMs is that frequently not all the available training examples are used in the description and specification of the separating hyperplane. The subset of points that lie on the margin (called support vectors) are the only ones that define the hyperplane of maximum margin.

The implementation of a linear SVM assumes that the multispectral feature data are linearly separable in the input space. In practice, data points of different class memberships (clusters) overlap one another. This makes linear separability difficult as the basic linear decision boundaries are often not sufficient to classify patterns with high accuracy. Techniques and workarounds such as the soft margin method (Cortes and Vapnik, 1995) and the kernel trick are used to solve the inseparability problem by introducing additional variables (called slack variables) in SVM optimization and mapping (using a suitable mathematical function) the nonlinear correlations into a higher (Euclidean or the Hilbert) space, respectively. A kernel function typically needs to fulfill Mercer's Theorem in order to be a valid kernel in SVMs (Scholkopf and Smola, 2001). The choice of a kernel function often has a bearing on the results of analysis. Furthermore, typical remote sensing problems usually involve identification of multiple classes (more than two). Adjustments are made to the simple SVM binary classifier to operate as a multi-class classifier using methods such as one-against-all, one-against-others, and directed acyclic graph (Knerr et al., 1990).

SVMs are particularly appealing in the remote sensing field due to their ability to successfully handle small training data sets, often producing higher classification accuracy than the traditional methods (Mantero et al., 2005). The underlying principle that benefits SVMs is the learning process that follows what is known as structural risk minimization. Under this scheme, SVMs minimize classification error on unseen data without prior assumptions made on the probability distribution of the data. Statistical techniques such as maximum likelihood estimation usually assume that data distribution is known a priori. Burges (1998) in a well-organized SVM tutorial described a simple experiment to illustrate an advantage of SVMs in an image recognition problem. In that demonstration, the performance of a basic multi-way SVM-based recognizer was assessed on image classification in the presence of prior knowledge. The accuracy turned out to be approximately the same if the pixels were first shuffled, with each image instance suffering the same random permutation. Yet, when the act of 'vandalism' (or removal of prior knowledge) took place, SVM still outperformed even the best neural networks. This discovery is particularly appealing in remote sensing applications since data acquired from remotely sensed imagery usually have unknown distributions, and methods such as Maximum Likelihood Estimation (MLE) that assume a multivariate normal data model do not necessarily match that assumption. Even if the data, whose dimensionality is assumed to match the number of spectral bands, were normally distributed, the assumption that the distribution can be described using a bell-shaped (Gaussian) function ceases to be sound, since the concentration of data in higher dimensional space tends to be in the tails (Fauvel et al., 2009). This phenomenon will continue to be encountered in remote sensing as new sensors increase spectral resolution and therefore data dimensionality.

There is also another interesting concept that serves as a key attraction to SVMs. Commonly described by many authors under the notion of overfitting (Montgomery and Peck, 1992), yet variously referred to by others as bias-variance tradeoff (Geman et al., 1992) or capacity control (Guyon et al., 1992), SVM-based classification has been known to strike the right balance between accuracy attained on a given finite amount of training patterns and the ability to generalize to unseen data.

Alongside the benefits derived from the SVM formulation there are also several challenges. The major setback concerning the applicability of SVMs is the choice of kernels. Although many options are available, some of the kernel functions may not provide optimal SVM configuration for remote sensing applications. Empirical evidence indicates that kernels such as radial basis function and polynomial kernels applied on SVM-based classification of satellite image data produce different results (Zhu and Blumberg, 2002). A good explanation on SVM kernels and their functionality is presented in numerous papers (e.g., Kavzoglu and Colkesen, 2009). From the non-expert user point of view, SVM theory is a bit intimidating, particularly due to the fact that the more efficient SVM variants often incorporate some difficult to understand concepts. This limits effective cross-disciplinary applications of SVMs.

Numerous SVM tutorials are available (such as Cortes and Vapnik (1995) and Burges (1998)), but none of these contains an exhaustive discussion on the increasing number of newly proposed variants of SVMs. In the remote sensing field a good starting point would be a textbook by Tso and Mather (2009) that provides a review of the entire field of classification methods for remotely sensed data, including SVMs. For those interested in rule extraction from SVMs a recent computer science review is available (Barakat