Remote Sensing of Environment 165 (2015) 135–147

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Image interpretation-guided supervised classification using nested segmentation

A.V. Egorov [a], M.C. Hansen [b,*], D.P. Roy [a], A. Kommareddy [b], P.V. Potapov [b]

[a] Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA
[b] University of Maryland, College Park, MD, USA

ARTICLE INFO

Article history:
Received 22 June 2014
Received in revised form 23 March 2015
Accepted 24 April 2015
Available online 23 May 2015

Keywords:
Land cover
Remote sensing
Classification
Active learning
Classifier
Landsat
Feature space partitioning
Nested segmentation

ABSTRACT

We present a new binary (two-class) supervised non-parametric classification approach that is based on iterative partitioning of multidimensional feature space into variably-sized and nested hyper-cubes (partitions). The proposed method contains elements of active learning and includes classifier to analyst queries. The spectral transition zone between two thematic classes (i.e., where training labels of different classes overlap in feature space) is targeted through iterative training derivation. Three partition categories are defined: *pure*, *indivisible* and *unlabeled*. *Pure* partitions contain training labels from only one class, *indivisible* partitions contain training data from different classes, and *unlabeled* partitions do not contain training data. A minimum spectral tolerance threshold defines the smallest partition volume to avoid over-fitting. In this way the transition zones between class distributions are minimized, thereby maximizing both the spectral volume of *pure* partitions in the feature space and the number of *pure* pixels in the classified image. The classification results are displayed to show each classified pixel's partition category (*pure*, *unlabeled* and *indivisible*). Mapping pixels belonging to *unlabeled* partitions serves as a query from the classifier to the analyst, targeting spectral regions absent of training data. The classification process is repeated until significant improvement of the classification is no longer realized or when no classification errors and *unlabeled* pixels are left. Variably-sized partitions lead to intensive training data derivation in the spectral transition zones between the target classes. The methodology is demonstrated for surface water and permanent snow and ice classifications using 30 m conterminous United States Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data time series from 2006 to 2010. The surface water result was compared with Shuttle Radar Topography Mission (SRTM) water body and National Land Cover Database (NLCD) open water classes with an overall agreement greater than 99% and Kappa coefficient greater than 0.9 in both of cases. In addition, the surface water result was compared with a classification generated using the same input data and a standard bagged Classification and Regression Tree (CART) classifier. The nested segmentation and CART-generated products had an overall agreement of 99.9 and Kappa coefficient of 0.99.

1. Introduction

Classification is regarded as a fundamental process in remote sensing used to relate pixel values to land cover or sometimes land use classes that are present at the corresponding location on the Earth's surface (Mather, 2004). Conventionally, pixel class assignment is determined by the spectral properties (signatures) of a given class or theme. Each spectral feature, for example red, near-infrared or shortwave infrared reflectance, is taken as an explanatory or independent variable. The theoretical *n*-dimensional space where *n* axes correspond to *n* raster bands in multispectral imagery, or *n* band transformations extracted from single images or time series, is often termed the feature space. Classifiers assign labels to pixels based on partitioning of feature space values using either unsupervised or training-based supervised methods.

Supervised classification methods have a long history since the development of techniques such as linear discriminant analysis (LDA) to classify two or more sub-populations (Fisher, 1936). Numerous classification algorithms have been developed and those applied to remotely sensed data include: k-nearest neighbor (kNN) (Fix & Hodges, 1951), multilayer perceptron (MLP) (Rosenblatt, 1957, 1958), maximum likelihood (ML) (Savage, 1976), Kohonen's self organized map (SOM) (Kohonen, 1982; Kohonen & Honkela, 2007), classification and regression trees (CART) (Breiman, Friedman, Olshen, & Stone, 1984), support vector machine (SVM) (Cortes & Vapnik, 1995), and random forests (RF) (Breiman, 2001). In supervised classification methods, training data of accurately labeled examples are taken as the dependent variable and associated to a set of independent variables. For land cover mapping using earth observation imagery, training data may be gathered on the basis of image interpretation, ground measurements or any other trusted source of information. In general, collecting training

* Corresponding author.

http://dx.doi.org/10.1016/j.rse.2015.04.022

data requires considerable time and effort. Supervised classification approaches are dependent on the experience of the remote sensing analyst in collecting training data and on the quality of the imagery. Supervised methods require *a priori* knowledge of the feature of investigation (e.g., the land cover type) in order to derive appropriate training data. Generating a training data set that accounts for all relevant spectral heterogeneity within and between classes is challenging and no systematic approach exists for training data collection. For example, training data selected by an analyst in the field may not be sufficiently representative of the conditions encountered in the image. Quality training data are required to achieve accurate supervised classification results.

Semi-automatic training set derivation has the goal of producing a parsimonious but sufficient set of training labels for supervised classification. Usually the acquisition of labeled data is difficult, time-consuming, or expensive to obtain. For these reasons a training set should be kept small while ensuring adequate classification performance. Several studies have shown however that classification accuracy increases with training set size (Lippitt, Rogan, Li, Eastman, & Jones, 2008; Rogan et al., 2008; Yan & Roy, 2015), although the optimal training size and distribution are usually unknown (Arora & Foody, 1997; Foody & Mathur, 2004b; Foody, McCulloch, & Yates, 1995; Pal & Mather, 2003; Zhuang, Engel, Lozanogarcia, Fernandez, & Johannsen, 1994). Many studies have emphasized the positioning of training data within the feature space, particularly the importance of collecting both pure (only one class in the pixel) and mixed pixel (more than one class in the pixel) training data. For example, Foody and Mathur (2004a,b, 2006) showed that the acquisition of training samples near feature space class boundaries may help reduce the training data set size without a loss of SVM classification accuracy. Similarly, Yu and Chi (2008) showed that a small training data set collected along class spectral boundaries provided comparable SVM classification accuracy to using training data consisting of a large number of pure pixels. Tuia, Pacifici, Kanevski, and Emery (2009) likewise employed a SVM and active learning to generate training data in classifying a series of single images. Other studies have shown similar results using mixed pixel training with aNN (Bernard, Wilkinson, & Kanellopoulos, 1997; Foody, 1999) and CART (Hansen, 2012) classifiers. Thus, a training set should be kept small, when training data collection is expensive, and should include both pure and mixed training data with particular emphasis on training data collection at the feature space class boundaries.

Semi-automatic training set derivation has been referred to as "active learning" in the machine learning literature and as "query learning" or "optimal experimental design" in the statistics literature (Settles, 2009). Active learning focuses on the interaction between the analyst (or some other information source) and the classifier. The model returns to the analyst the pixels whose classification outcome is the most uncertain. After accurate labeling by the analyst, pixels are added to the training set in order to reinforce the model. In this way, the model is optimized on well-chosen difficult examples, maximizing its generalization capabilities (Tuia, Volpi, Copa, Kanevski, & Munoz-Mari, 2011). Semi-automatic learning can be of great practical value in many real-word problems where unlabeled data are abundant or easily obtained, but the acquisition of labeled data is difficult, time-consuming, or expensive to obtain (Lippitt et al., 2008; Settles, 2009). Active learning algorithms have been studied in many real world problems, such as classifying handwritten characters (Lang & Baum, 1992), part-of-speech tagging (Dagan & Engelson, 1995), sensor scheduling (Krishnamurthy, 2002), learning ranking functions for information retrieval (Yu, 2005), word sense disambiguation (Fujii, Tokunaga, Inui, & Tanaka, 1998), text classification (Hoi, Jin, & Lyu, 2006; Lewis & Catlett, 1994; McCallum & Nigam, 1998; Tong & Koller, 2000), information extraction (Settles & Craven, 2008; Thompson, Califf, & Mooney, 1999), video classification and retrieval (Hauptmann, Lin, Yan, Yang, & Chen, 2006; Yan, Yang, & Hauptmann, 2003), speech recognition (Tür et al., 2005), and cancer diagnosis (Liu, 2004). Active learning is also suitable for remote sensing applications, where the number of pixels among which the search is performed is large and manual definition is redundant and time consuming. However, only a relatively few studies have been dedicated to remote sensing data classification using active learning (e.g. Jackson & Landgrebe, 2001; Jun & Ghosh, 2008; Li, Bioucas-Dias, & Plaza, 2010; Licciardi et al., 2009; Tuia et al., 2009, 2011).

This study builds on previous research by presenting a semi-automatic active learning classification approach called *nested segmentation*. Nested segmentation identifies areas in need of labeling followed by manual assignment by an analyst. The resulting systematic feature space partitioning defines the classification rules, i.e., unlike other active learning classification approaches (Tuia et al., 2009) an extant classification algorithm is not used. The approach is iterated until either a preset classification accuracy is acquired or there are no unlabeled classified pixels. Instead of relying simply on the size of the training data set to produce a quality classification, we focus on two other training set properties, *representativeness* and *concentration*. Training data that sufficiently cover the intra-class spectral variation per land cover type are representative. Training data that are densely located along spectral class boundaries are concentrated. Training data representativeness is achieved by identifying and adding training data in regions of the feature space that lack training samples. Training data concentration is achieved by identifying regions of the feature space where different classes overlap, targeting the addition of training data and recursively sub-dividing the particular spectral region. This allows the analyst's efforts to be focused on deriving training where more intensive sampling is needed. The method provides a new way of iteratively collecting training data for a binary classification that allows an analyst to collect a compact and sufficient training data set.

The nested segmentation approach is designed to be fast in its implementation and appropriate for large area mapping tasks at national to global scales that normally require large training data sets. Mapping at such scales presents a challenge for training data set derivation due to the variety of intra- and interclass spectral variation present. For example, at national scales, surface water can range from clearly identifiable low turbidity lakes to more challenging water bodies, including turbulent coastal surface waters and briny inland lakes of endorheic basins. Land covers such as dark conifer forests or central business districts featuring tall buildings can be confused with open water bodies. The presented method is meant to target all such variations in a rapid, iterative fashion. The methodology is first described and then demonstrated by application to 5 years of 30 m conterminous United States Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Web Enabled Landsat (WELD) data (Roy et al., 2010) to generate open surface water (SW) and permanent snow and ice (SI) classifications. The SW classification is compared quantitatively with water masks from the Shuttle Radar Topography Mission (SRTM) water body data set (Rabus, Eineder, Roth, & Balmer, 2003) and the National Land Cover Database (NLCD2006) open water class (Fry et al., 2011). In addition, the WELD nested segmentation SW classification is compared with a SW classification generated from the same training and Landsat data but using a standard bagged CART classifier. This is followed by a brief discussion of the methodology and implications for future research.

2. Data and pre-processing

2.1. Landsat data

The Landsat satellite series, operated by the U.S. Department of Interior/U.S. Geological Survey (USGS) Landsat project, with satellite development and launches engineered by the National Aeronautics and Space Administration (NASA), represent the longest dedicated land remote sensing data record (Roy, Wulder, et al., 2014). Landsat data provide a balance between requirements for localized moderate spatial resolution studies and global monitoring (Goward, Masek, Williams, Irons, & Thompson, 2001). Free of charge radiometrically