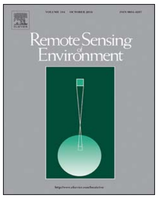
ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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



Mapping forest change using stacked generalization: An ensemble approach

Sean P. Healey[a,*], Warren B. Cohen[b], Zhiqiang Yang[c], C. Kenneth Brewer[d], Evan B. Brooks[e], Noel Gorelick[f], Alexander J. Hernandez[g], Chengquan Huang[h], M. Joseph Hughes[c], Robert E. Kennedy[c], Thomas R. Loveland[i], Gretchen G. Moisen[a], Todd A. Schroeder[j], Stephen V. Stehman[k], James E. Vogelmann[i], Curtis E. Woodcock[l], Limin Yang[i], Zhe Zhu[m]

[a] *US Forest Service Rocky Mountain Research Station, United States*
[b] *US Forest Service Pacific Northwest Research Station, United States*
[c] *Oregon State University, United States*
[d] *US Forest Service Washington Office, United States*
[e] *Virginia Polytechnic Institute and State University, United States*
[f] *Google Corporation, United States*
[g] *Utah State University, United States*
[h] *University of Maryland, United States*
[i] *USGS Earth Resources Observations and Science Center, United States*
[j] *US Forest Service Southern Research Station, United States*
[k] *State University of New York College of Environmental Science and Forestry, United States*
[l] *Boston University, United States*
[m] *Texas Tech University, United States*

ARTICLE INFO

Keywords:
Multiple classifier systems
Landsat
Change detection
Ensemble

ABSTRACT

The ever-increasing volume and accessibility of remote sensing data has spawned many alternative approaches for mapping important environmental features and processes. For example, there are several viable but highly varied strategies for using time series of Landsat imagery to detect changes in forest cover. Performance among algorithms varies across complex natural systems, and it is reasonable to ask if aggregating the strengths of an ensemble of classifiers might result in increased overall accuracy. Relatively simple rules have been used in the past to aggregate classifications among remotely sensed maps (e.g. using majority predictions), and in other fields, empirical models have been used to create situationally specific algorithm weights. The latter process, called "stacked generalization" (or "stacking"), typically uses a parametric model for the fusion of algorithm outputs. We tested the performance of several leading forest disturbance detection algorithms against ensembles of the outputs of those same algorithms based upon stacking using both parametric and Random Forests-based fusion rules. Stacking using a Random Forests model cut omission and commission error rates in half in many cases in relation to individual change detection algorithms, and cut error rates by one quarter compared to more conventional parametric stacking. Stacking also offers two auxiliary benefits: alignment of outputs to the precise definitions built into a particular set of empirical calibration data; and, outputs which may be adjusted such that map class totals match independent estimates of change in each year. In general, ensemble predictions improve when new inputs are added that are both informative and uncorrelated with existing ensemble components. As increased use of cloud-based computing makes ensemble mapping methods more accessible, the most useful new algorithms may be those that specialize in providing spectral, temporal, or thematic information not already available through members of existing ensembles.

1. Introduction

1.1. The challenge of mapping subtle forest cover loss

Land cover change due to both human and natural disturbance processes has a profound effect on how ecosystems function, affecting biogeochemical (Chambers et al., 2007; Kurz et al., 2009) and hydrological cycles (Seilheimer et al., 2013), habitat conditions (Spies et al., 2010), and availability of social and economic human benefits (González-Olabarria and Pukkala, 2011). Characterization of land cover

---

* Corresponding author.
E-mail address: seanhealey@fs.fed.us (S.P. Healey).

http://dx.doi.org/10.1016/j.rse.2017.09.029
Received 9 March 2017; Received in revised form 16 September 2017; Accepted 22 September 2017
0034-4257/ Published by Elsevier Inc.

Please cite this article as: Healey, S.P., Remote Sensing of Environment (2017), http://dx.doi.org/10.1016/j.rse.2017.09.029

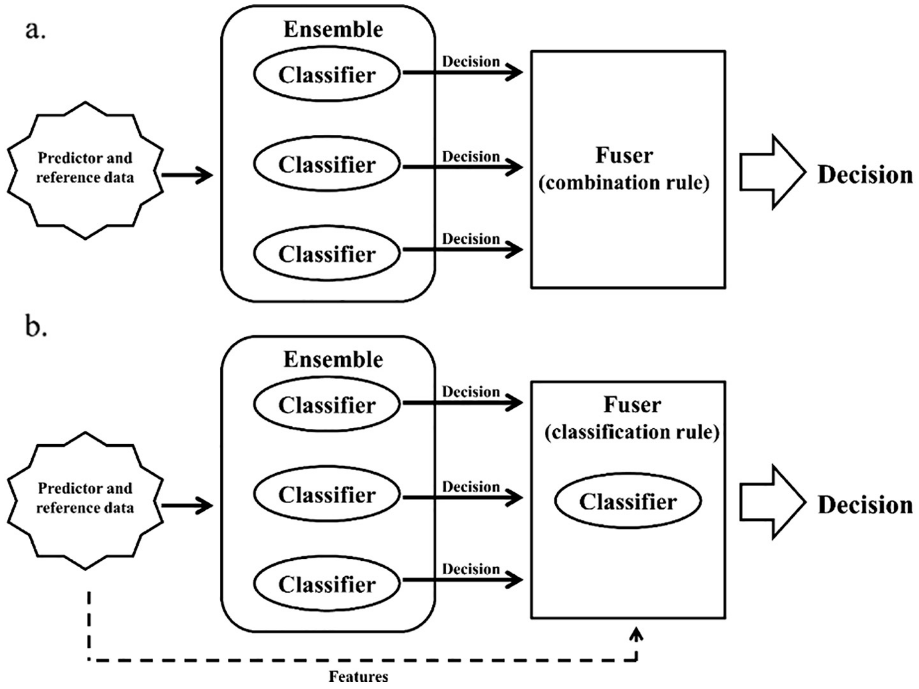


Fig. 1. Topologies of commonly used MCS. Fuser functions may either use a combination rule, such as voting or averaging, which requires only classifier outputs (panel a), or they may call upon features from a learning dataset to facilitate weighting of outputs on the basis of model performance (panel b). This second approach has been termed "stacking."
Figure adapted from Woźniak et al. (2014).

change has therefore emerged as a discipline with a central bearing on many fields of study (Turner et al., 2007). The Landsat platform has been a primary source of land change information, capable of detecting important vegetative and disturbance patterns because of the sensor's long history and appropriate temporal, spatial, and spectral properties (Cohen and Goward, 2004). The Sentinel and SPOT platforms have also proven useful for this task (e.g., (Antropov et al., 2016; Li et al., 2016; Verhegghen et al., 2016)). The free release of all images in the Landsat archive (in 2008) has led to the development of many new algorithms capable of using temporally dense observations to increase the breadth, accuracy, and precision of land cover change characteristics that can be mapped (Wulder et al., 2012).

However, like most remote sensing problems, there are many factors that can increase the complexity of detecting forest change, particularly beyond the relatively straightforward stand-replacing disturbances targeted in earlier efforts (e.g. Healey et al., 2008). Cohen et al.'s (2016) national survey of forest disturbance processes found that low-magnitude forest decline was the most common cause of disturbance, particularly in the Western US. Likewise, US Forest Service inventory data indicates that partial harvests are more commonly practiced than clearcuts across the country (Smith et al., 2009), and the inter-agency Monitoring Trends in Burn Severity project (Schwind et al., 2010) found that only 36% of the area burned by 13,400 large fires in the US between 1985 and 2010 had moderate or greater severity (Finco et al., 2012). For any given low-magnitude disturbance, subtle removals of forest canopy may increase spectral reflectance in both the visible and mid-infrared wavelengths if removal of vegetation reveals brighter soils, but reflectance may actually decrease if canopy removal increases the contribution of shadowing to the spectral signal or if charring occurs (Schroeder et al., 2011). Consistency of spectral response across space and time may also be compromised by phenology, atmosphere, topography, soil type, forest type, and forest structure.

There are several change detection algorithms which target lower-magnitude change (e.g. (DeVries et al., 2015; Healey et al., 2006; Meigs et al., 2015)) in very specific scenarios, but it is an open question if Landsat or other remote sensing platforms can be used across complex landscapes to detect the full range of disturbance magnitudes and types without also introducing detrimental levels of false-positive (i.e., commission) error. It should be noted that while the term "change detection" is used here for the process of mapping forest disturbance, that process is very much subject to error and actually represents a prediction of change more than a definitive discovery. The more accurate "change prediction" is not used here both because of convention and to distinguish the current monitoring task from work involved with assessment of future events (e.g. (Seidl et al., 2014).

1.2. Multiple classifier systems

This paper presents a test of the idea that an ensemble of change detection algorithms can be used together to obtain forest disturbance maps of greater accuracy and sensitivity than maps from any single automated algorithm. Wolpert and Macready (1997) demonstrated, in their "No Free Lunch" theorems, that if an algorithm performs well in one class of problems, it necessarily "pays" for that accuracy with degraded performance on a set of all remaining problems. If different algorithms have different specialties, particularly if those specialties are diverse, combination of those algorithms in Multiple Classifier Systems (MCS) should improve global performance (Oza and Tumer, 2008). We use the term "classifier" to refer to any generalizing algorithm or model that produces a hypothesis about an object using a set of learning data. A variety of tools have been used as classifiers across disciplines, from logistic regression to nearest neighbor imputation and support vector machine methods (e.g. (Sáez et al., 2013); (Kavzoglu et al., 2014)), and this paper focuses on a variety of algorithms that make use of time series analysis with Landsat imagery.

Analytical approaches based on MCS now play an important role in tasks ranging from detecting computer security risks to diagnosing disease (Woźniak et al., 2014). This paper focuses on a class of MCS which applies an ensemble of classifiers to a problem simultaneously and then uses a fusion rule to employ a meta-classification process.

Fig. 1 illustrates two types of fusion rules: one which uses a simple combination rubric such as an average or majority (a), and one which uses a secondary model to re-weight the classifiers according to their performance against similar cases in the reference data (b). Random Forests (RF; (Breiman, 2001)) is a prominent example of an MCS which uses a combination rule. RF creates an ensemble of similar classifiers by training decision tree-based models with random partitions of the input data, a process called "bagging" (Breiman, 1996a). An RF prediction is