ISPRS Journal of Photogrammetry and Remote Sensing 120 (2016) 53–64

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Mapping raised bogs with an iterative one-class classification approach

Benjamin Mack[a,b,*], Ribana Roscher[a,c], Stefanie Stenzel[b], Hannes Feilhauer[d], Sebastian Schmidtlein[b], Björn Waske[a]

[a] Institute of Geographical Sciences - Remote Sensing and Geoinformatics, Freie Universität Berlin, Malteserstraße 74-100, 12249 Berlin, Germany
[b] Institute of Geography and Geoecology, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany
[c] Institute of Geodesy and Geoinformation, University of Bonn, Nußallee 15, 53115 Bonn, Germany
[d] Institute of Geography, FAU Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, Germany

ARTICLE INFO

Article history:
Received 6 September 2015
Received in revised form 19 June 2016
Accepted 25 July 2016
Available online 19 September 2016

Keywords:
Remote sensing
Land cover classification
RapidEye
Natura 2000
Biased Support Vector Machine
MAXENT

ABSTRACT

Land use and land cover maps are one of the most commonly used remote sensing products. In many applications the user only requires a map of one particular class of interest, e.g. a specific vegetation type or an invasive species. One-class classifiers are appealing alternatives to common supervised classifiers because they can be trained with labeled training data of the class of interest only. However, training an accurate one-class classification (OCC) model is challenging, particularly when facing a large image, a small class and few training samples. To tackle these problems we propose an iterative OCC approach. The presented approach uses a biased Support Vector Machine as core classifier. In an iterative pre-classification step a large part of the pixels not belonging to the class of interest is classified. The remaining data is classified by a final classifier with a novel model and threshold selection approach. The specific objective of our study is the classification of raised bogs in a study site in southeast Germany, using multi-seasonal RapidEye data and a small number of training sample. Results demonstrate that the iterative OCC outperforms other state of the art one-class classifiers and approaches for model selection. The study highlights the potential of the proposed approach for an efficient and improved mapping of small classes such as raised bogs. Overall the proposed approach constitutes a feasible approach and useful modification of a regular one-class classifier.

1. Introduction

Rapid environmental change threatens valuable habitats around the world. Monitoring the conservation status of habitats is important for environmental governance and management. In the European Union the Natura 2000 network, which comprises sites designated under the Habitats Directive (Council of the European Communities, 1992) and the Birds Directive (Council of the European Communities, 2010), aims on the protection of such habitats. It comprises more than 26000 sites covering 950000 km^2, or 17.5%, of the terrestrial area of the EU (Evans et al., 2013). Every six years, the member states are obliged to report on the status of the habitats in and outside the Natura 2000 network. Efficient tools and strategies are required for a cost-effective monitoring of these large areas. It is widely accepted that remote sensing has the potential to assist in achieving this goal (Corbane et al., 2015; Vanden Borre et al., 2011) and this potential is currently evaluated by various research activities. In this context, mapping the distribution and the delineation change of vegetation types (Schuster et al., 2015; Stenzel et al., 2014; Corbane et al., 2013; Alexandridis et al., 2009) is as important as mapping their state (Neumann et al., 2015; Zlinszky et al., 2015; Möckel et al., 2014; Schuster et al., 2011). Vegetation type mapping by the classification of remote sensing data makes use of the distinctive spectral reflectance characteristics at a particular time of the year or over the year. Different pilot studies have proved that a variety of vegetation types can be classified with remote sensing imagery (see Corbane et al., 2015 for an exhaustive review).

Monitoring vegetation types in the complete territory constitutes another classification problem which requires different solutions (Stenzel et al., 2014). The problem is to map one or a few classes of interest in a large area which is mainly covered by numerous classes which do not need to be separated. Conventional supervised classifiers need to be trained with a representative set of samples covering an exhaustive set of classes and are, thus, inefficient for classifying one or few classes of interest (Foody

* Corresponding author at: Institute of Geographical Sciences - Remote Sensing and Geoinformatics, Freie Universität Berlin, Malteserstraße 74-100, 12249 Berlin, Germany.

E-mail address: benjamin.mack@fu-berlin.de (B. Mack).

http://dx.doi.org/10.1016/j.isprsjprs.2016.07.008
0924-2716/

et al., 2006). Classification approaches that do not require an exhaustive training set can be found under the terms one-class classification (OCC), partially supervised classification and classification with reject option. In this work we concentrate on OCC which has already been used successfully for mapping vegetation types relevant for Natura 2000 with remotely sensed data.

The one-class Support Vector Machines OCSVM (Schölkopf et al., 2001) and the similar Support Vector Data Description (SVDD) (Tax and Duin, 2004) have been used successfully, e.g., for the classification of saltmarshes (Sanchez-Hernandez et al., 2007a) and fenland (Sanchez-Hernandez et al., 2007b) using Landsat data, and for mapping a specific tree specie (*Tabebuia guayacan*) using high resolution Quickbird data (Sánchez-Azofeifa et al., 2011). However, several comparative studies (Baldeck and Asner, 2015; Muñoz Marí et al., 2010; Li and Guo, 2010) revealed the poorer performance of the OCSVM compared to other one-class classifiers such as the biased SVM (BSVM) (Liu et al., 2003), MAXENT (Phillips and Dudík, 2008; Elith et al., 2011) and the PUL algorithm (Li et al., 2011). These approaches were used successfully for classifying raised bogs and other Natura 2000 habitat types (Stenzel et al., 2014), different heath and shrub formations (Morán-Ordóñez et al., 2012) specific (tree) species (Baldeck and Asner, 2015; Evangelista et al., 2009) and several other land use and land cover types (Wan et al., 2015; Lin et al., 2014; Ortiz et al., 2013; Li et al., 2011; Muñoz Marí et al., 2010). Recently, also sparse representation-based classifier has become common to solve classification tasks with only one class of interest. They assume that each data point can be modeled by a weighted sparse linear combination of basis elements collected in a dictionary. Based on the reconstruction error a data point can be identified whether it belongs to the class of interest (e.g., Song et al., 2016).

The above mentioned one-class classifiers can be distinguished dependent on the input data used for training: P-classifiers are only trained on labeled data of the positive (P) class (i.e. the classes of interest), while PU-classifiers additionally learn from unlabeled (U) data. Prominent examples of the first group are OCSVM and SVDD while BSVM and MAXENT belong to the second group. The superiority of BSVM and MAXENT compared to OCSVM (Baldeck and Asner, 2015; Muñoz Marí et al., 2010; Li and Guo, 2010) can be attributed to the additional information that can be extracted from unlabeled data. However, usually the PU-classifiers are computationally more expensive (see below) (Muñoz Marí et al., 2010).

Model selection (eventually including threshold selection) is a critical step during the training stage of flexible and thus versatile classifiers and is particularly challenging in the case of OCC. Since reference data is not available for the ''other'' (or negative) class, a performance criteria (PC) for model selection can only be derived from P-(Muñoz Marí et al., 2010) or PU-data (Li and Liu, 2003) (see Section 3.2). However, particularly with limited positive training data model selection becomes challenging and state-of-the-art approaches might fail. This may result in an inadequate model selection and thus, limited classification accuracies (Baldeck and Asner, 2015; Mack et al., 2014).

MAXENT has an important advantage compared to most other one-class classifiers where model selection is a crucial step to be implemented by the user. The freely available (for research activities) and easy to use software implementation (https://www.cs.princeton.edu/schapire/maxent/) comes with a default parameterization that has been proven to perform well in all the mentioned studies and thus release the user from implementing model selection based on PU-data. This might explain why MAXENT is used more frequently than other one-class classifiers in applied studies. However, for deriving a binary classification, a threshold has to be applied to the continuous output (often called suitabilities) returned by the MAXENT software. This is a major difficulty and to the best of our knowledge there is no fully automatic method available for selecting an optimal threshold. Instead, most (if not all) studies applying MAXENT for land cover/land use classification with remote sensing data use subjective expert decisions (Li and Guo, 2010; Ortiz et al., 2013) or apply empirical models based on data of the ''other'' class (Evangelista et al., 2009; Lin et al., 2014).

With respect to PU-classifiers, the benefit in terms of accuracy comes with the burden of higher computational cost (Muñoz Marí et al., 2010) which may increase strongly with the number of training samples. Thus, the question arises how large the set of unlabeled training samples should be. If the subset is too large processing can become computationally too expensive or even unfeasible. On the other hand, a small subset may not contain the relevant information required for deriving an accurate model and therefore may result in low mapping accuracies. Iterative OCC approaches (Yu, 2005) are particularly useful when the data to be classified is large and only a small part of if belongs to the class of interest. In such situations iterative approaches can outperform state-of-the-art classifiers in terms of accuracy and computational cost by concentrating on more difficult pixels.

In this paper we present an iterative classification approach to accurately and computational efficiently map raised bogs with positive and unlabeled data in an area south of Munich, Germany. For the study site, covering approximately 20 km $\times$ 50 km, multi-seasonal RapidEye data is available for classification. The approach is particularly designed for such classification problems where the occurrence of the class of interest is very small as is the number of positive training samples.

The proposed approach can be separated in two main steps: pre-classification and final classification. During pre-classification a computationally efficient one-class classifier is applied iteratively in order to classify pixels which very likely belong to the negative class. After convergence, the final classifier is applied onto the remaining pixels. It is important to note that the final one-class classifier operates on a subset of the image which contains the positive pixels and the part of the negative pixels which are most similar to the positive labeled training data. The balancedness of the positive and negative classes in this subset is much higher compared to the complete image. As a consequence, a relative small sample from this data should be sufficient for the PU-classifier in order to extract the relevant information from the unlabeled data. Furthermore, model and threshold selection methods can be applied which would be inefficient, unreliable or even unfeasible when used with extremely unbalanced data (Sezgin and Sankur, 2004). In the final classification a novel approach based on a normal mixture model is used for model and threshold selection. In the proposed approach, a biased Support Vector Machine (BSVM) (Liu et al., 2003) is used as core classifier and is referred to as iterative BSVM (iBSVM).

The major novelties of the presented approach can be summarized as follows: (i) The iterative pre-classification approach is designed in a way that all but an insignificant amount of the positive samples and a similar amount of negative class pixels remain in the resulting subset, thus resulting in a balanced subset. (ii) The approach has been modified with respect to previous approaches (Yu, 2005) such that over-iteration, i.e. rejecting a significant amount of positive pixels, is impossible. (iii) The integrated model and threshold selection approach is a novel contribution which significantly differs from the few existing approaches (Muñoz Marí et al., 2010; Li and Liu, 2003). It is worth to stress that the usage of the sophisticated model selection approach in the final classification would not be feasible without the pre-classification.

In order to show the effectiveness of the proposed approach it is compared to the state-of-the-art one-class classifiers OCSVM, BSVM and MAXENT. In case of the OCSVM and the BSVM the two state-of-the-art model selection approaches are analyzed in detail. These approaches are (i) well known in the fields of machine learning, pattern recognition, and/or applied remote