ISPRS Journal of Photogrammetry and Remote Sensing 136 (2018) 1–12

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Comparing the performance of flat and hierarchical Habitat/Land-Cover classification models in a NATURA 2000 site

Yoni Gavish [a,*], Jerome O'Connell [b], Charles J. Marsh [a], Cristina Tarantino [c], Palma Blonda [c], Valeria Tomaselli [d], William E. Kunin [a]

[a] School of Biology, University of Leeds, Leeds LS2 9JT, United Kingdom
[b] School of Biosystems and Food Engineering, University College Dublin, D04 N2E5, Ireland
[c] Institute of Atmospheric Pollution Research (IIA), National Research Council (CNR), c/o Interateneo Physics Department, Via Amendola 173, 70126 Bari, Italy
[d] Institute of Biosciences and BioResources (IBBR), National Research Council (CNR-IBBR), via G.Amendola 165/A, 70126 Bari, Italy

ARTICLE INFO

Article history:
Received 25 May 2017
Received in revised form 19 September 2017
Accepted 4 December 2017

Keywords:
Classification
Machine-learning
Hierarchical models
Random forest
NATURA 2000
Habitat/Land-Cover

ABSTRACT

The increasing need for high quality Habitat/Land-Cover (H/LC) maps has triggered considerable research into novel machine-learning based classification models. In many cases, H/LC classes follow pre-defined hierarchical classification schemes (e.g., CORINE), in which fine H/LC categories are thematically nested within more general categories. However, none of the existing machine-learning algorithms account for this pre-defined hierarchical structure. Here we introduce a novel Random Forest (RF) based application of hierarchical classification, which fits a separate local classification model in every branching point of the thematic tree, and then integrates all the different local models to a single global prediction. We applied the hierarchal RF approach in a NATURA 2000 site in Italy, using two land-cover (CORINE, FAO-LCCS) and one habitat classification scheme (EUNIS) that differ from one another in the shape of the class hierarchy. For all 3 classification schemes, both the hierarchical model and a flat model alternative provided accurate predictions, with kappa values mostly above 0.9 (despite using only 2.2–3.2% of the study area as training cells). The flat approach slightly outperformed the hierarchical models when the hierarchy was relatively simple, while the hierarchical model worked better under more complex thematic hierarchies. Most misclassifications came from habitat pairs that are thematically distant yet spectrally similar. In 2 out of 3 classification schemes, the additional constraints of the hierarchical model resulted with fewer such serious misclassifications relative to the flat model. The hierarchical model also provided valuable information on variable importance which can shed light into "black-box" based machine learning algorithms like RF. We suggest various ways by which hierarchical classification models can increase the accuracy and interpretability of H/LC classification maps.

1. Introduction

Human-mediated changes in the distribution of habitats and land-cover types are one of the main drivers of the global biodiversity crisis. Consequently, providing reliable Habitat/Land-Cover (H/LC) maps for various conservation related issues is of high priority. For example, H/LC maps are used as input layers for species distribution models (Carlson et al., 2014; Coops et al., 2016; Thuiller et al., 2004) or as obligatory background layers for conservation of umbrella species with well-defined habitat requirements (Li and Pimm, 2016; Murphy and Noon, 1992). Furthermore, H/LC maps are fundamental for mapping ecosystem services (e.g., Koschke et al., 2012) and for natural capital assessments (Brown et al., 2016). Finally, in many cases, the habitats themselves are targeted for conservation and management. For example, as part of the EU Habitat Directive (EU, 2007), all member states of the European Union are required to periodically produce H/LC maps and use the maps for change detection and conservation status assessment. Hence further developing our ability to produce H/LC

Abbreviations: H/LC, Habitat/Land-Cover; RF, Random Forest; FRF, Flat Random Forest; HRF, Hierarchical Random Forest; OoB, Out of Bag; H.Step, hierarchical stepwise majority rule; H.Mult, hierarchical multiplicative majority rule; FAO-LCCS, Food and Agriculture Organisation's Land Cover Classification System; EUNIS, EUropean Nature Information System habitat classification system; Hie.F, Hierarchical F measure.

* Corresponding author at: School of biology, University of Leeds, Leeds LS2 9JT, United Kingdom.
E-mail addresses: gavishyoni@gmail.com (Y. Gavish), jerome.oconnell@ucd.ie (J. O'Connell), charliem2003@gmail.com (C.J. Marsh), cristina.tarantino@iia.cnr.it (C. Tarantino), palma.blonda@iia.cnr.it (P. Blonda), valeria.tomaselli@ibbr.cnr.it (V. Tomaselli), W.E.Kunin@leeds.ac.uk (W.E. Kunin).

https://doi.org/10.1016/j.isprsjprs.2017.12.002
0924-2716/

maps at fine thematic and spatial resolution over wide extents is essential for effective conservation, planning, monitoring, reporting and management of natural resources. As a consequence, there has been a recent surge of methodological and conceptual developments in the field of H/LC classification (Blaschke, 2010; Corbane et al., 2015; Lu and Weng, 2007; Lucas et al., 2015, 2011; Myint et al., 2011; Tso and Mather, 2009; Xie et al., 2008).

In recent years the usage of machine-learning algorithms has become increasingly popular (Belgiu and Drăguţ, 2016) as these machine-learning algorithms are efficient at identifying complex classification rule sets, thus potentially providing accurate classification outputs with relatively little investment of time and effort. In many cases, the H/LC classified by machine-learning algorithms rely on a pre-defined national or international classification schemes (e.g., CORINE), to allow a common language of communication between scientists, management agents and policy makers. Most classification schemes adopt a hierarchical, tree-like structure due to several advantages of such structures. Firstly, classes within a hierarchical classification scheme can be grouped into more abstract classes based on semantic similarity criteria, i.e., a hierarchical H/LC class set comprises several semantic granularities. Secondly, a hierarchical H/LC class set can be applied to a variety of spatial scales (each spatial scale requiring the selection of a scale-specific semantic granularity). The former characteristic is particularly useful to meet the minimum required accuracy standard when a specific subclass accuracy is below this standard (Congalton, 1991) and/or when it is difficult to differentiate between subclasses at a given spatial scale. For example, the European Nature Information System habitat classification scheme (EUNIS) has a tree-like structure with up to eight hierarchical levels, containing a total of 5282 habitat classes (at all levels). CORINE LC has three hierarchical levels, with a total of 44 LC classes. Similarly, classification schemes invented ad-hoc for more local studies may also have a hierarchical structure (e.g., Haest et al., 2017). However, most machine learning algorithms follow a flat classification approach (sensu Silla and Freitas, 2011) in which all H/LCs are classified simultaneously in a 'one-against-all' approach. In other words, machine learning algorithms ignore information on the thematic hierarchy that forms the conceptual basis of most classification schemes. Interestingly, many knowledge-based classifiers follow a top-down approach, in which experts first provide rules (e.g., spectral) that separate general H/LC classes from one another, and then move down the thematic tree while providing more specific rules for more specific H/LC categories (e.g., Lucas et al., 2011, 2007).

There are several reasons why incorporating such hierarchical information into the analytical pathway may be beneficial. First, the rule-sets produced by most machine learning algorithms are a 'black-box' to the users because of their size and complexity. It is therefore very difficult to understand or visualise what variables are important in distinguishing between specific sets of habitats. A hierarchical approach may shed some light into the 'black-box' by providing information on variable importance in various locations along the class hierarchy. Second, habitats that are thematically close to one another are not necessarily ecologically/spectrally similar. For example, a forest and grassland may both be listed under the thematic group of 'non-crop' habitats while a wheat field will occur under the thematic group of 'crops' habitats. However, ecologically and spectrally, the grassland may resemble the wheat field more than the forest. A flat classification approach ignores the thematic proximity altogether, while a hierarchical approach will first invest considerable effort in distinguishing 'crop' from 'non-crop', thus potentially preventing confusion between spectrally similar yet thematically distant habitats. Third, if the number of habitats is large, the flat approach may not be able to deal with the complexity of the thematic data, while a hierarchical approach could break the problem into manageable portions by partitioning the feature-space of each group into lower dimensions.

Finally, it has been shown that incorporating the hierarchical structure into the modelling framework can increase model accuracy (Thoonen et al., 2013). More specifically, Silla and Freitas (2011) found that various hierarchical approaches tended to increase model accuracy in a wide range of classification problems, especially when misclassifications are weighted by their distance along the classification hierarchy (Kiritchenko et al., 2005). Such hierarchical measures of accuracy acknowledge that not all misclassifications are as critical as the others, e.g., misclassifying one broadleaved-woodland habitat as an alternative closely-related woodland type is arguably a less critical mistake than misclassifying it as a grassland or saltmarsh. In addition, flat classification models only provide performance measures for the entire model or at the H/LC level (i.e., user and producer accuracies). Hierarchical classification models provide the same information with additional accuracy for each local model. That is, the hierarchical approach also provides accuracy for sets of H/LCs that share a common ancestor along the class hierarchy. This information may be crucial for decision makers that may be less interested in the overall accuracy of a map and more by its ability to provide reliable information on sets of H/LCs they care most about (e.g., how well does this model classify non-crop habitats?).

We are aware of only a few published manuscripts that focused on hierarchical, machine-learning based classification methods in the remote-sensing literature. Melgani and Bruzzone (2004) found that several support-vector-machine based hierarchical models outperformed flat models when classifying 9 land-use classes in northwest Indiana. Thoonen et al. (2013) found that a tree-structure Markov random field (TS-MRF) method, which captures the hierarchical thematic structure as well as contextual information, outperformed flat classification methods for heathland areas in Belgium. O'Connell et al. (2015) accounted for spatial hierarchy (nested objects) and thematic hierarchy (2 levels). They reported slightly better classification outcomes (compared to a flat approach) when the probabilities from a Random Forest (RF; Breiman, 2001) model trained at the top level of the thematic hierarchy where included as predictors of RF models trained at the lower level of the thematic hierarchy. Pena et al. (2014) compared flat and hierarchical approaches (based on 4 different algorithms) for mapping cropland areas and found that the flat approach was slightly outperformed by a support-vector-machine based hierarchical model, which fitted a local classifier per parent node. They also found the hierarchical approach increased the minimum sensitivity at the crop level. Finally, Haest et al. (2017) applied an hierarchical classification along four thematic levels when classifying heathland vegetation types for conservation status assessment. They followed a top-down approach such that the class selected for a given pixel in level 2 of the hierarchy could only be one of the children classes of the class selected in level 1 (with similar rules for levels 3 and 4). Haest et al. (2017) observed higher accuracies for the hierarchical approach compared to a flat approach.

In this paper we introduce a novel application of hierarchical classification based on the RF algorithm, which accounts for the pre-defined hierarchical structure of classification schemes. The application is available for use in a new R package, entitled '*HieRanFor*'. We tested the hierarchical approach in a NATURA 2000 study site from Italy, using three different classification schemes. Our main aim is to compare the performance of the hierarchical and flat approaches and to explore if the variables identified as important at various locations along the class hierarchy provide meaningful ecological knowledge of the system.