Information Sciences 324 (2015) 60–87

Contents lists available at ScienceDirect

Information Sciences

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

Improving land cover classification through contextual-based optimum-path forest

D. Osaku [a], R.Y.M. Nakamura [b], L.A.M. Pereira [c], R.J. Pisani [d], A.L.M. Levada [a], F.A.M. Cappabianco [e], A.X. Falcão [c], João P. Papa [f,*]

[a] Department of Computing, UFSCar—Univ Federal of São Carlos, São Carlos, Brazil
[b] Big Data Brasil, São Paulo, Brazil
[c] Institute of Computing, University of Campinas, Campinas, Brazil
[d] Western University of São Paulo, Presidente Prudente, SP, Brazil
[e] Federal University of São Paulo, São José dos Campos, Brazil
[f] Department of Computing, São Paulo State University, Bauru, Brazil

ARTICLE INFO

Article history:
Received 14 January 2015
Revised 14 May 2015
Accepted 12 June 2015
Available online 2 July 2015

Keywords:
Land cover classification
Optimum-path forest
Contextual classification

ABSTRACT

Traditional machine learning algorithms very often assume statistically independent data samples. However, this is clearly not the case in remote sensing image applications, in which pixels present spatial and/or temporal dependencies. In this work, it has been presented an approach to improve land cover image classification using a contextual approach based on optimum-path forest (OPF) and the well-known Markov random fields (MRFs), hereinafter called OPF–MRF. In addition, it is also introduced a framework to the optimization of the amount of contextual information used by OPF–MRF. Experiments over high- and medium-resolution satellite (CBERS-2B, Landsat 5 TM, Ikonos-2 MS and Geoeye) and radar (ALOS-PALSAR) images covering the area of two Brazilian cities have shown the proposed approach can overcome several shortcomings related to standard OPF classification. In some cases, the proposed approach outperformed traditional OPF in about 9% of recognition rate, which is crucial for land cover classification.

1. Introduction

Pixel-based classification plays an important role in many remote sensing applications, such as land cover and target recognition, just to name a few. Remote sensing image classification is often performed by support vector machines—SVMs [3,19,21], artificial neural networks—ANNs [17,19], and optimum-path forest—OPF [28,35], but without taking into account the spatial and/or temporal dependencies among pixels. However, these approaches can generate undesirable artefacts (e.g., salt-and-pepper effect) as reported in some works [8,38]. Contextual classifiers have been proposed to address the problem by exploiting the spatial dependency of nearby pixels that are likely to belong to a same class [5].

Tarabalka et al. [40], for instance, proposed a contextual approach based on support vector machines and Markov random fields (MRFs) for remote sensing image classification called SVM–MRF. Moser and Serpico [27] also proposed a version of the

* Corresponding author. Tel.: +55 14 3103 6079; fax: +55 14 3103 6079.
E-mail addresses: danosaku@hotmail.com, alexandre@dc.ufscar.br (D. Osaku), rodrigo.mizobe@bigdata.inf.br (R.Y.M. Nakamura), luis.pereira@ic.unicamp.br (L.A.M. Pereira), pisanigeo@gmail.com (R.J. Pisani), fcappabianco@gmail.com (F.A.M. Cappabianco), afalcao@ic.unicamp.br (A.X. Falcão), papa@fc.unesp.br, papa.joaopaulo@gmail.com (J.P. Papa).

http://dx.doi.org/10.1016/j.ins.2015.06.020

SVM–MRF approach using a single-step formulation, rather than the two-phase strategy presented in [40]. Kittler and Föglein [20] combined a Bayesian classifier and a Gaussian MRF by using pixel spectral and spatial dependencies. Shekhar et al. [38] employed a Bayesian classifier with different models to capture spatial context. They provided comparisons among spatial autoregression and two different Markov models. Another related work was proposed by Zhang et al. [44], which used SVMs with an adaptive MRF-based approach (a-MRF) to avoid overcorrection on the boundary between classes. Other approaches for contextual classification of remote sensing images exploit hierarchical and multi-resolution feature description [7,9,41]. Additionally, Frery et al. [12] performed multispectral image classification by means of Gaussian maximum likelihood and contextual iterated conditional modes (ICM) algorithm.

In the context of land cover classification using contextual information, the reader can find some interesting works as well. Laha et al. [23], for instance, applied evidence theory to incorporate contextual information for further image classification using fuzzy rules. Ghimire et al. [15] applied random forests, and Stuckens et al. [39] employed a linkage-based clustering algorithm for land cover classification using contextual information. Sarkar et al. [37] presented a hybrid approach among MRF and Dempster–Shafer theory. It is also valuable to shed light over the seminal work of Wharton [42], which presented a contextual land cover classification approach based on the local distribution of nearby labels, as well as the work of Guo et al. [16], which proposed a contextual approach based on cascaded classifiers for remote sensing imagery classification. Recently, a comprehensive review about contextual-based classifiers has been presented by Li et al. [24]. Finally, Mahmoudi et al. [26] stated the contextual relations can overcome some challenges in urban areas recognition based on satellite imagery. Although Aghighi et al. [1,2] have also presented an approach to estimate a smoothing parameter that controls the amount of interactions between the spatial and contextual information based on dynamic blocks, SVM and class label co-occurrence matrices, their works differ from this one, since here we employed meta-heuristic techniques to address the problem of contextual-based classification concerning the OPF classifier.

In short, we propose a meta-heuristic-based optimization framework to find suitable values for the parameter that controls the amount of contextual information used in the classification process. Additionally, we introduce a contextual classifier named OPF–MRF based on MRF and optimum-path forest for supervised learning [32,33]. The traditional OPF classifier interprets training samples (feature vectors) as nodes of a complete graph, identifies key samples (prototypes) in all classes, and partitions the graph into an optimum-path forest rooted at those prototypes. The classification of a new sample is quickly performed in an incremental way by finding its most closely connected training sample and assigning its class to it. Such a strategy has demonstrated to be similar/superior to SVMs and ANNs in several applications, being much faster for training, since this version of OPF is parameter-independent.

In the OPF–MRF approach, the output of OPF classifier is interpreted as an MRF and improved by ICM algorithm [6] in order to add contextual (sample labels) to the original feature vectors. A second OPF classifier, trained with the extended feature vectors, outputs the final label map. As two independent and preliminary works, the OPF–MRF classifier was previously presented for magnetic resonance image classification [29], and the proposed meta-heuristic-based framework was previously evaluated in the same context as well [31]. In this paper, it is the first time we combine both approaches for contextual classification and evaluate them for land cover classification. The work have shown one can improve vanilla OPF classification in the context of land cover recognition using images obtained by four satellites with different spatial resolutions, as well as one radar image. This is important to highlight the proposed approach can obtain better results than pixel-based OPF classification in distinct scenarios.

The remainder of this paper is organized as follows. Sections 2 and 3 describe the OPF theory background, as well as its contextual version, respectively. The experimental setup and experiments are presented in Sections 4 and 5, respectively. Section 6 states conclusions and future works.

2. Optimum-path forest classification

The OPF framework is a recent highlight to the development of pattern recognition techniques based on graph partitions. The nodes are the data samples, which are represented by their corresponding feature vectors, and are connected according to some predefined adjacency relation. Given some key nodes (prototypes), they will compete among themselves aiming at conquering the remaining nodes. Thus, the algorithm outputs an optimum path forest, which is a collection of optimum-path trees (OPTs) rooted at each prototype. This work employs the OPF classifier proposed by Papa et al. [32,33], which is explained in more details as follows.

Let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ be a labeled dataset, such that $\mathcal{D}_1$ and $\mathcal{D}_2$ stands for the training and test sets, respectively. Let $\mathcal{S} \subset \mathcal{D}_1$ be a set of prototypes of all classes (i.e., key samples that best represent the classes). Let $(\mathcal{D}_1, A)$ be a complete graph whose nodes are the samples in $\mathcal{D}_1$, and any pair of samples defines an arc in $A = \mathcal{D}_1 \times \mathcal{D}_1$ (Fig. 1a).[1] Additionally, let π_s be a path in $(\mathcal{D}_1, A)$ with terminus at sample $s \in D_1$.

The OPF algorithm proposed by Papa et al. [32,33] employs the path-cost function $f_{\max}$ due to its theoretical properties for estimating prototypes (Section 2.1 gives further details about this procedure):

$$f_{\max}(\langle s \rangle) = \begin{cases} 0 & \text{if } s \in S \\ +\infty & \text{otherwise,} \end{cases}$$
$$f_{\max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{\max}(\pi_s), d(s,t)\}, \tag{1}$$

[1] The arcs are weighted by the distance between their corresponding nodes.