Pattern Recognition Letters 83 (2016) 188–194

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A new approach to contextual learning using interval arithmetic and its applications for land-use classification☆

Danillo Roberto Pereira[1], João Paulo Papa*

São Paulo State University, Department of Computing, Bauru, SP, Brazil

ARTICLE INFO

Article history:
Available online 14 April 2016

Keywords:
Sliding Window
Sequential learning
Contextual learning
Interval Arithmetic

ABSTRACT

Contextual-based classification has been paramount in the last years, since spatial and temporal information play an important role during the process of learning the behavior of the data. Sequential learning is also often employed in this context in order to augment the feature vector of a given sample with information about its neighborhood. However, most part of works describe the samples using features obtained through standard arithmetic tools, which may not reflect the data as a whole. In this work, we introduced the Interval Arithmetic to the context of land-use classification in satellite images by describing a given sample and its neighbors using interval of values, thus allowing a better representation of the model. Experiments over four satellite images using two distinct supervised classifiers showed we can considerably improve sequential learning-oriented pattern classification using concepts from Interval Arithmetic.

1. Introduction

Machine learning techniques have become a page-turner in the way we organize and analyze data that come from different areas, ranging from engineering to medicine and economics. Although traditional pattern recognition techniques very often consider the samples are independent to each other, there are many other applications that do not fit in such models, such as time series in finance-related problems and meteorological observations, just to name a few. In some cases, the nature of the problem suggests a temporal ordering of the data, e.g., audio and speech processing, as stated by Ryabko [20]. In other applications, the ordering may be only tangentially related to time, as in natural language processing, or even completely unrelated to temporal notion (analysis of biological sequences). Therefore, considering such a priori knowledge may lead us to more accurate learners.

In the context of image classification, a way to introduce priori knowledge in the problem formulation is to use smoothness constraints in order to consider the spatial context of the data. When looking at a picture or a video, we can clearly see the pixels vary smoothly in homogeneous regions. Sequential- and contextual-oriented learning are some well-known methodologies very often used to address situations in which spatial and/or contextual information may help the classifier into better modeling the behavior of the data, as stated by Cohen and Carvalho [2], which introduced the Stacked Sequential Learning (SSL), as well as by Kittler and Föglein] [9] and Dietterich [3], that presented an interesting review about sequential learning techniques. The authors also highlighted the high computational load of some techniques based on such idea. Later on, Gatta et al. [8] proposed a multi-scale sequential learning approach (Multi-scale Stacked Sequential Learning with Muti-resolution decomposition — MSSL-MR, and Multi-scale Stacked Sequential Learning with Pyramid decomposition — MSSL-PY), in which the contextual information is obtained not only from the sample's neighborhood, but also from pixels farther away. The idea of multiple scales is driven by several Gaussian-convolved labeled images, which are former obtained by means of a traditional classification process. Afterwards, Puertas et al. [19] addressed the aforementioned work in the context of multi-class-based classification problems. Sampedro et al. [21] proposed a similar approach to that one introduced by Puertas et al. [19], but now in a three-dimensional space, which has been used together with error-correcting output codes in the context of medical image classification.

An interesting approach is to design hybrid versions of well-known classifiers in order to consider contextual information by means of Markov Random Fields (MRFs). Osaku et al. [14], for instance, proposed the OPF-MRF, which is a contextual version of the Optimum-Path Forest (OPF) classifier, and Tarabalka et al. [22]

☆ This paper has been recommended for acceptance by Jenny Qian Du Eckart Michaelsen.

* Corresponding author. Tel.: +55 1431036079; fax: +55 1431036079.
E-mail addresses: danilopereira@unoeste.br (D.R. Pereira), papa@fc.unesp.br, papa.joaopaulo@gmail.com (J.P. Papa).
[1] Tel.: +55-18-99703-2445.

http://dx.doi.org/10.1016/j.patrec.2016.03.020
0167-8655/

presented the Support Vector Machines (SVM) classifier integrated with Markov Random Fields (SVM-MRF). Both works have addressed remote sensing-oriented applications. Fauvel et al. [7] proposed a contextual approach based on spectral and spatial information for the classification of high-resolution remote sensing images, and Wehmann and Liu [23] applied contextual classification by means of Markovian kernels aiming at change detection in satellite images. Very recently, Pereira et al. [18] evaluated the OPF classifier in the context of sequential learning for land-use satellite image classification, achieving more accurate results than naïve OPF.

Despite all the good results of such approaches, most part of them rely on extracting some information over a neighborhood of a given sample, for further pattern classification. Such techniques usually employ feature vectors based on scalar values that belong to standard arithmetic tools. In this work, we introduce the concept of Interval Arithmetic (IA) proposed by Moore [11] in the context of sequential learning-based pattern recognition. The Interval Arithmetic represents a scalar number in a finite interval of values, thus generalizing the standard arithmetic. Interval Arithmetic concepts were widely applied in the fuzzy set theory to address image processing and recognition problems [6,10,13,15]. Alefeld and Mayer [1] also presented an interesting review about some IA-based applications, as well as its theoretical background.

However, the reader can find very few works that employed IA to the context of machine learning applications. Drago and Ridella [4], for instance, employed IA together with single Perceptron networks, and the very same group of authors validated Interval Arithmetic in the context Multilayer Perceptron neural networks (please, refer to the work conducted by Drago and Ridella [5]). In this paper, we have shown how to achieve more accurate results in land-use classification by using features obtained through Interval Arithmetic concepts. We validated the proposed approach against with the Sliding Window (SW) technique, which is very usual in the sequential learning research field, as well as against with the aforementioned SSL, MSSL-MR and MSSL-PY. Experiments over four satellites images and two distinct classifiers showed the proposed approach can obtain much more accurate results than using representations based on standard arithmetic. In short, this paper has two main contributions: (i) to introduce IA in the context of sequential learning, and (ii) to propose a new sequential learning technique based on SW and multi-resolution decomposition. The remainder of this work is organized as follows. Section 2 and 3 present the theoretical background about the Interval Arithmetic and Sliding Window methodologies, respectively. The proposed approach to apply IA concepts in sequential learning is discussed in Section 4, and Section 5 and 6 present the methodology and experiments, respectively. Finally, Section 7 states conclusions and future works.

2. Interval Arithmetic

The Interval Arithmetic was proposed by Moore in the 1960's, being defined as a range-based computation model where each interval $[x]$ is represented by a non-empty real-valued range $[x_l, x_h]$ that encodes the subset of real numbers r that satisfy the following condition:

$$w = \{r \in \mathbb{R}^* / x_l \leq r \leq x_h\}. \tag{1}$$

The IA background theory defines a set of relations and operations over the intervals [10,12]. Therefore, we can compare, join, sum and even multiply intervals of numbers, being a more powerful tool than traditional arithmetic, since any real number r can be represented by the singular interval $[r, r]$. Besides, the Interval Arithmetic is a useful apparatus to provide efficient representations of error bounds and uncertainty. Below, we present the definition of the main operations regarding intervals:

Intersection. The intersection between two intervals $[x]$ and $[y]$ is defined as follows:

$$[x] \cap [y] = [\max\{x_l, y_l\}, \min\{x_h, y_h\}], \tag{2}$$

being defined only when $\max\{x_l, y_l\} \leq \min\{x_h, y_h\}$.

Union. The union operation between two intervals $[x]$ and $[y]$ is only defined for intervals that do not present empty intersection, i.e., $[x] \cap [y] \neq \emptyset$:

$$[x] \cup [y] = [\min\{x_l, y_l\}, \max\{x_h, y_h\}]. \tag{3}$$

Convex Hull. The convex hull of two intervals $[x]$ and $[y]$ is the smallest interval that contains both intervals, i.e.:

$$[x] \breve{\cup} [y] = [\min\{x_l, y_l\}, \max\{x_h, y_h\}]. \tag{4}$$

Although the above formulation is the very same presented in Eq. 3, it does not require the empty intersection constraint.

The IA also defines comparison primitives, as follows:

Equality. $[x] = [y] \iff x_l = y_l \text{ and } x_h = y_h$

Lower than:. $[x] < [y] \iff x_h < y_l$

Lower than or equal:. $[x] \leq [y] \iff x_h \leq y_l$

Greater than:. $[x] > [y] \iff x_l > y_h$

Greater than or equal:. $[x] \geq [y] \iff x_l \geq y_h$

The basic arithmetic operations were also extended to intervals, as follows:

Summation. $[x] + [y] = [x_l + y_l, x_h + y_h]$

Negation. $-[x] = [-x_h, -x_l]$

Subtraction. $[x] - [y] = [x] + (-[y]) = [x_l - y_h, x_h - y_l]$

Multiplication. $[x].[y] = [\min\{x_l y_l, x_l y_h, x_h y_l, x_h y_h\}, \max\{x_l y_l, x_l y_h, x_h y_l, x_h y_h\}]$

3. Sequential learning with Sliding Window

The Sliding Window is an approach often used for the pattern classification using concepts of sequential learning, as stated by Dietterich [3]. The main idea is to model the contextual information by means of features obtained from the neighborhood of a given sample. Since this approach is very usual and well-known in the specific community, we opted to dedicate a separated section to it, which will serve as a basis to the explanation of the proposed approach.

In pixel-based image classification, a straightforward and naïve representation of each pixel (x_i, y_i) from an image I can be achieved by using its brightness values, i.e., $I(x_i, y_i)$, $i = 1, 2, \ldots, M$, being M the number of pixels of that image. However, SW employs the neighborhood features in order to design an extended feature vector for that pixel. Therefore, given an image I and a square neighborhood window $\mathcal{W}_{h \times h}$ of size $h \times h$, each pixel $I(x_i, y_i)$ is represented by the brightness of the pixels that fall in its neighborhood, which is defined by the window $\mathcal{W}_{h \times h}$ centered at that pixel (x_i, y_i). For the sake of explanation, consider the pixel (x, y) and a window

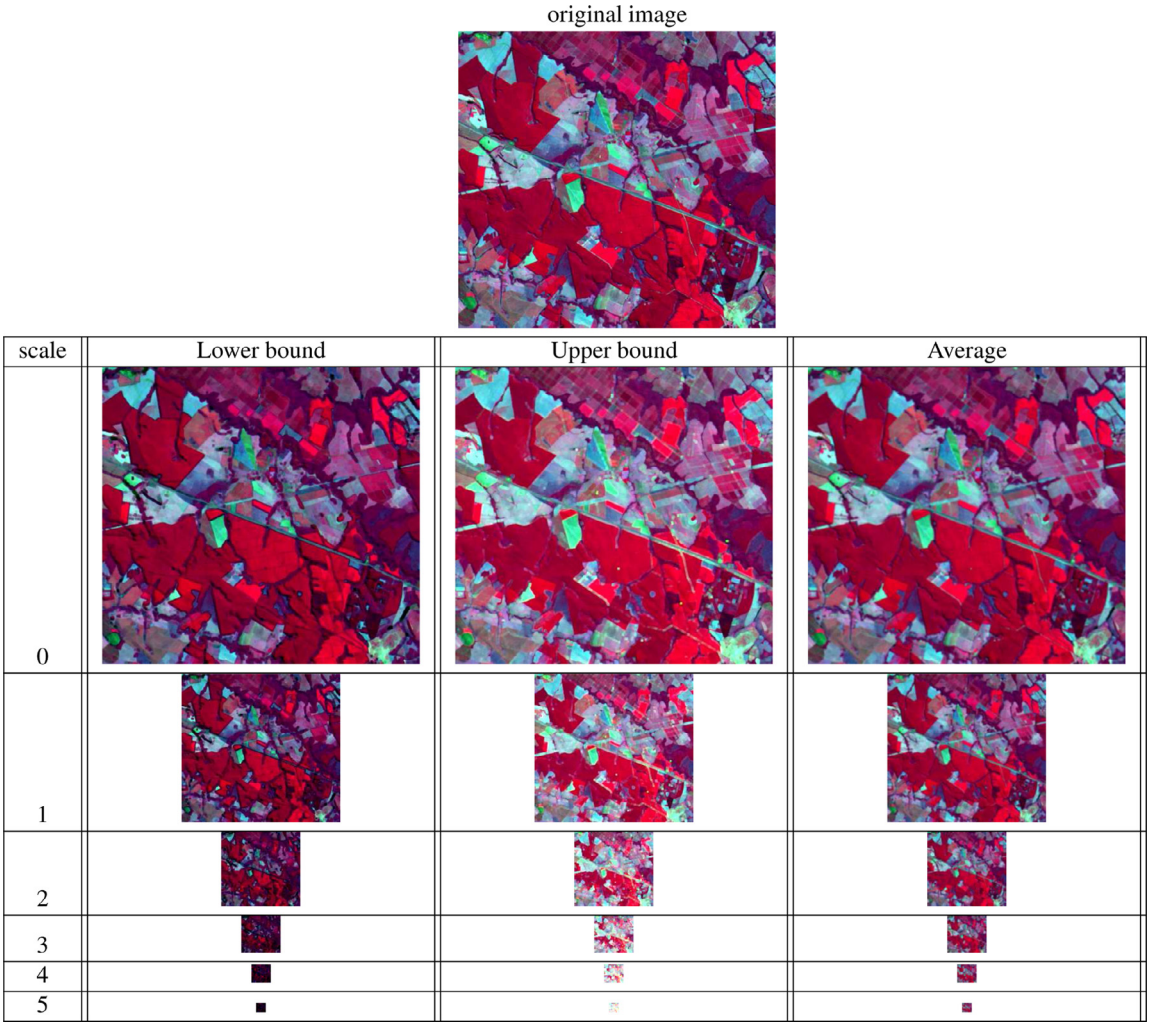

Fig. 1. The image pyramid decomposition for the lower and upper bounds, as well as the average value of the interval.

neighborhood $\mathcal{W}_{3\times3} = \{a, b, c, d, e, f, g, h\}$. The feature vector $\mathbf{f}_{(x,y)}$ of that pixel is then defined as follows:

$$\mathbf{f}_{(x,y)} = \{a, b, c, d, (x, y), e, f, g, h\}, \tag{5}$$

where the set of pixels $\mathcal{W}_{3\times3}$ obeys a clockwise orientation.

However, it is worth pointing out two important facts about the Sliding Window approach that motivated us to use Interval Arithmetic: (i) the feature extension process used by SW is sensitive to any sort of rotation or exchanging of pixels that may occur inside the window, and (ii) the SW can generate high-dimensional representations, which may not be desirable when taking into account the computational load of some techniques. The next section shows how to avoid such shortcomings with the tools provided by Interval Arithmetic.

4. Interval Arithmetic applied for the classification of remote sensing images

In this section, we present the proposed approach to handle the problem of land-use image classification by means of Interval Arithmetic, hereinafter called Interval Arithmetic-based Contextual Learning (IA-CL). The first step is to model the image space using a multi-scale representation as follows: given an input image I with $nx \times ny$ pixels and a reduction factor of α, we build S versions of the image I, i.e., $I^{(1)}$, ..., $I^{(S)}$, where $S = S^{max} - 1$, and $S^{max} = log_{\alpha}(\min\{nx, ny\})$. Notice the image at scale $I^{(S)}$ has about 1×1 of resolution, and $I^{(0)}$ stands for the original image.

As such, we model the task of image description using a pyramid-based representation, in such a way the set of features is defined over larger regions in order to exploit their contextual (neighborhood) information, which is not available using a pixel-wise-driven representation. The hierarchy built up over different scales enables us to exploit interactions among distinct image representations, thus providing more accurate descriptions for the further classification of land-use regions.

As aforementioned, the idea of the proposed approach is to generate a more robust representation than the one given by Eq. 5. Therefore, the new representation for each pixel is obtained as follows: given a pixel $(x, y) \in I^{(0)}$, we first place a square window of size $(2\alpha^i + 1) \times (2\alpha^i + 1)$ centered at that pixel, and all pixel's brightness that fall in that window are encoded as an interval of values. We then compute the lower and upper bounds of that interval through the convex hull formulation (Eq. 4), as well as we compute the average value of the interval. After that, the very same procedure is applied for each $I^{(i)}$, $i \in \{1, 2, \ldots, S\}$, and the values computed at each scale are then concatenated[2] into a single feature vector for the pixel $(x, y) \in I^{(0)}$. Regardless the scale,

[2] Notice the window is always centered at pixel $(x, y) \in I^{(0)}$, even for different scales $I^{(i)}, i \in \{1, 2, \ldots, S\}$.

the patches are always computed with respect to the original image, thus being centered at the pixel $(x, y) \in I^{(0)}$.

It is worth noting to mention the user can control the step size concerning the sampling process. Therefore, if the step is smaller than the window size, we shall have a certain amount of overlapping among adjacent windows. In our test we use purposeful overlapping to exploit the superposition of the patches. The proposed decomposition creates three pyramid representations, being two for the lower and upper bounds, and one more for the average value. The images at scale i that encode the lower and upper bounds, and the average values are denoted by $I_{lo}^{(i)}$, $I_{hi}^{(i)}$ and $I_{avg}^{(i)}$, respectively. The pyramids of the lower and upper bounds are composed of images that encode the uncertainty of the pixels within that neighborhood; and the average image represents an additional information about the behavior of the pixels that belong to that interval. Fig. 1 displays an example of the proposed decomposition process with $S = 5$ scales, and the images generated at each scale.

The whole process can be summarized as follows: after the generation of the pyramid decomposition, we obtain an extended feature vector for each pixel composed of $S+1$ intervals (the lower and upper bounds) and $S+1$ scalars (the average value of the interval at each scale). The interval at scale i is represented by $r_i = [I_{lo}^{(i)}(x\alpha^i, y\alpha^i), I_{hi}^{(i)}(x\alpha^i, y\alpha^i)]$, and the average value at the very same scale i is represented by $I_{avg}^{(i)}(x\alpha^i, y\alpha^i)$. Although we have considered other primitives from Interval Arithmetic basis (e.g., intersection, sum and multiplication), the convex hull was the one that provided the best results.

Therefore, the extended feature vector $\mathbf{f}^*_{(x,y)}$ for each pixel $(x, y) \in I$ is then represented by the following features:

$$\mathbf{f}^*_{(x,y)} = \{ I(x,y), \left[I_{lo}^{(0)}(\tfrac{x}{\alpha^0}, \tfrac{y}{\alpha^0}), I_{hi}^{(0)}(\tfrac{x}{\alpha^0}, \tfrac{y}{\alpha^0})\right], I_{avg}^{(0)}(\tfrac{x}{\alpha^0}, \tfrac{y}{\alpha^0}), \ldots, \left[I_{lo}^{(S)}(\tfrac{x}{\alpha^S}, \tfrac{y}{\alpha^S}), I_{hi}^{(S)}(\tfrac{x}{\alpha^S}, \tfrac{y}{\alpha^S})\right], I_{avg}^{(S)}(\tfrac{x}{\alpha^S}, \tfrac{y}{\alpha^S}) \}$$

Differently from other techniques, such as SW and the ones based on SSL, IA-CL is not sensitive to rotations or permutations of the pixels inside the window, since it considers the lower and upper bounds given by the convex hull, as well as the average value inside that window.

5. Methodology

In this section, we present the methodology employed to validate IA-ML in the context of land-cover image classification using two distinct supervised pattern recognition techniques: the Optimum-Path Forest (OPF) classifier, which was firstly proposed by Papa et al. [17] and later on enhanced by Papa et al. [16], and the well-known Bayesian classifier, hereinafter called Bayes. Since both techniques are parameterless and have a low computational burden, they seemed to be a good choice for the purpose of this paper. However, we would like to emphasize the proposed approach can be used with any other pattern recognition technique.

In regard to the land-cover data, we used images obtained from CBERS-2B and Landsat 5 TM covering the area of Itatinga, SP-Brazil, and other images were obtained from Ikonos-2 MS and Geoeye covering the area of Duque de Caxias, RJ-Brazil. Fig. 2 displays these images, being their respective ground truth versions illustrated in Fig. 3.

We compared IA-CL against five approaches (standard classification, SW, SSL, MSSL-MR and MSSL-PY) under distinct scenarios: we evaluated the influence of different training set sizes with 5%, 10% and 20% of the entire image, being the remaining pixels used

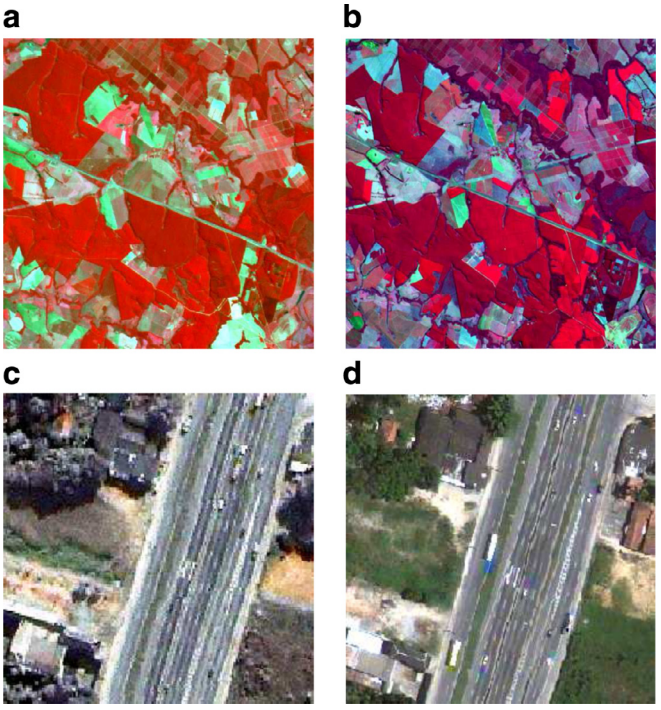

Fig. 2. Satellite images used in the experiments: covering the area of Itatinga, SP — Brazil by (a) CBERS-2B CCD (20 m) sensor (R2G3B4) and (b) Landsat 5 TM (30 m) sensor (R4G3B5), and covering the area of Duque de Caxias, RJ — Brazil by (c) Ikonos-2 MS sensor (R4G3B2) and (d) Geoeye sensor (R5G4B3). The CBERS-2B and Landsat 5 TM images have 526 × 492 pixels, and Ikonos-2 MS and Geoeye images have 258 × 250 and 268 × 250 pixels, respectively. Notice that Ikonos-2 MS and Geoeye images were obtained through a fusion process between the corresponding images from MS (4 m) and PAN (1 m) sensors using the pan-sharpening method. The final image has a spatial resolution of 1 m.

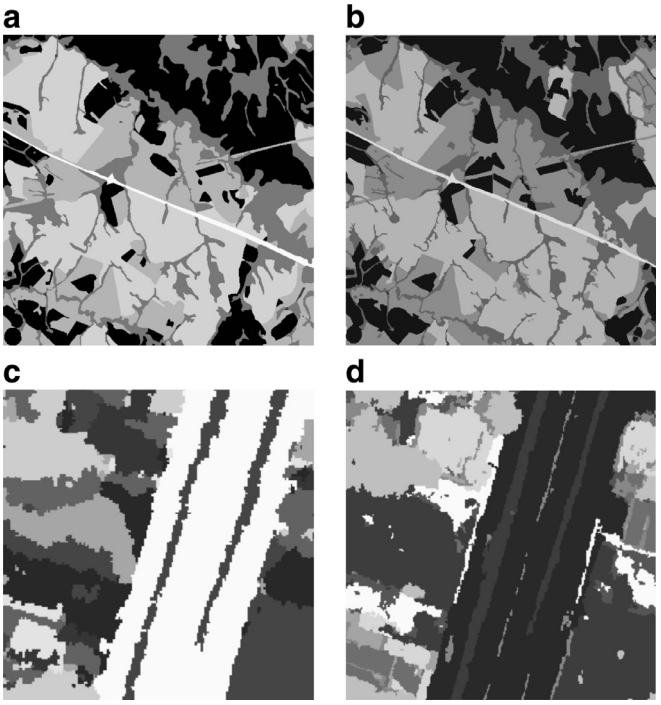

Fig. 3. Labeled images used in the experiments: (a) and (b) refer to the images displayed in Fig. 2(a) and (b), respectively, and (c) and (d) stand for images displayed in Fig. 2(c) and (d), respectively.

Table 1
Experimental results concerning CBERS-2B image using 5%, 10% and 20% of the entire image for training purposes.

	Accuracy (5%)	Accuracy (10%)	Accuracy (20%)
OPF	67.9 ± 0.1	68.1 ± 0.0	67.5 ± 0.0
OPF-SW	78.5 ± 0.1	81.9 ± 0.1	83.8 ± 0.0
OPF-SSL	61.5 ± 0.2	61.7 ± 0.1	60.3 ± 0.1
OPF-MSSL-MR	64.3 ± 0.0	65.4 ± 0.1	67.2 ± 0.1
OPF-MSSL-PY	62.1 ± 0.2	63.2 ± 0.0	63.2 ± 0.0
OPF-CA-SL (proposed)	**89.9 ± 0.4**	**92.2 ± 0.1**	**93.7 ± 0.0**
Bayes	72.0 ± 0.0	72.2 ± 0.1	71.0 ± 0.0
Bayes-SW	79.2 ± 0.0	81.1 ± 0.0	84.0 ± 0.0
Bayes-SSL	66.5 ± 0.0	66.9 ± 0.0	68.0 ± 0.0
Bayes-MSSL-MR	64.7 ± 0.0	66.9 ± 0.1	68.0 ± 0.0
Bayes-MSSL-PY	64.9 ± 0.1	66.8 ± 0.1	68.3 ± 0.2
Bayes-CA-SL (proposed)	**90.0 ± 0.9**	**92.0 ± 0.1**	**94.1 ± 0.0**

Table 2
Experimental results concerning Geoeye image using 5%, 10% and 20% of the entire image for training purposes.

	Accuracy (5%)	Accuracy (10%)	Accuracy (20%)
OPF	71.0 ± 0.1	72.6 ± 0.5	73.4 ± 0.2
OPF-SW	76.1 ± 0.2	77.1 ± 0.1	77.8 ± 0.2
OPF-SSL	66.4 ± 0.1	67.1 ± 0.0	69.1 ± 0.0
OPF-MSSL-MR	64.6 ± 0.4	65.9 ± 0.3	66.2 ± 0.1
OPF-MSSL-PY	64.9 ± 0.1	65.7 ± 0.1	68.0 ± 0.0
OPF-IA-CL (proposed)	**86.2 ± 0.0**	**88.8 ± 0.0**	**90.9 ± 0.1**
Bayes	70.3 ± 0.0	70.8 ± 0.0	71.6 ± 0.1
Bayes-SW	75.6 ± 0.1	77.0 ± 0.0	77.9 ± 0.1
Bayes-SSL	69.0 ± 0.0	70.1 ± 0.0	72.4 ± 0.3
Bayes-MSSL-MR	65.6 ± 0.1	67.1 ± 0.0	68.9 ± 0.2
Bayes-MSSL-PY	66.6 ± 0.1	68.0 ± 0.0	69.1 ± 0.0
Bayes-IA-CL (proposed)	**86.5 ± 0.0**	**89.2 ± 0.1**	**91.2 ± 0.0**

Table 3
Experimental results concerning Ikonos image using 5%, 10% and 20% of the entire image for training purposes.

	Accuracy (5%)	Accuracy (10%)	Accuracy (20%)
OPF	69.2 ± 0.1	71.3 ± 0.2	74.3 ± 0.1
OPF-SW	70.4 ± 0.0	71.5 ± 0.0	72.9 ± 0.0
OPF-SSL	61.2 ± 0.2	62.0 ± 0.1	67.8 ± 0.2
OPF-MSSL-MR	62.4 ± 0.3	64.2 ± 0.2	68.5 ± 0.1
OPF-MSSL-PY	62.4 ± 0.3	63.0 ± 0.0	67.2 ± 0.4
OPF-IA-CL (proposed)	**84.6 ± 0.0**	**87.5 ± 0.1**	**90.2 ± 0.1**
Bayes	69.0 ± 0.9	70.1 ± 0.1	73.8 ± 0.1
Bayes-SW	70.7 ± 0.2	71.8 ± 0.0	72.8 ± 0.1
Bayes-SSL	67.2 ± 0.2	65.2 ± 0.1	70.1 ± 0.2
Bayes-MSSL-MR	60.1 ± 0.9	64.2 ± 0.2	68.0 ± 0.0
Bayes-MSSL-PY	61.0 ± 0.1	63.9 ± 0.0	69.1 ± 0.0
Bayes-IA-CL (proposed)	**84.7 ± 0.1**	**87.8 ± 0.1**	**90.8 ± 0.0**

to compose the test set. In order to allow a robust statistical evaluation, we performed a cross-validation procedure with 15 runnings for further computing the Wilcoxon signed-rank test [24]. Additionally, each pixel has been described by its RGB values to compose the dataset samples for pattern classification purposes[3].

6. Experiments

In this section, we present the experimental results regarding the proposed approach. Table 1, 2, 3 and 4 present the mean accuracy results considering CBERS-2B, Landsat 5 TM, Ikonos-2 MS and Geoeye images, respectively[4]. The most accurate techniques considering the Wilcoxon signed-rank test are highlighted in bold.

Table 4
Experimental results concerning Landsat image using 5%, 10% and 20% of the entire image for training purposes.

	Accuracy (5%)	Accuracy (10%)	Accuracy (20%)
OPF	64.8 ± 0.2	64.9 ± 0.0	66.0 ± 0.3
OPF-SW	83.6 ± 0.2	84.6 ± 0.2	86.6 ± 0.1
OPF-SSL	69.2 ± 0.1	70.0 ± 0.0	71.3 ± 0.1
OPF-MSSL-MR	69.2 ± 0.1	69.7 ± 0.0	71.1 ± 0.1
OPF-MSSL-PY	69.2 ± 0.1	69.0 ± 0.0	70.1 ± 0.2
OPF-IA-CL (proposed)	**85.7 ± 0.1**	**86.9 ± 0.1**	**88.4 ± 0.1**
Bayes	69.6 ± 0.1	70.4 ± 0.0	69.0 ± 0.2
Bayes-SW	83.6 ± 0.2	86.3 ± 0.2	87.0 ± 0.1
Bayes-SSL	72.9 ± 0.0	73.4 ± 0.1	74.5 ± 0.3
Bayes-MSSL-MR	70.6 ± 0.1	71.4 ± 0.0	73.0 ± 0.0
Bayes-MSSL-PY	71.3 ± 0.1	72.5 ± 0.2	73.2 ± 0.0
Bayes-IA-CL (proposed)	**85.4 ± 0.1**	**86.9 ± 0.1**	**88.8 ± 0.1**

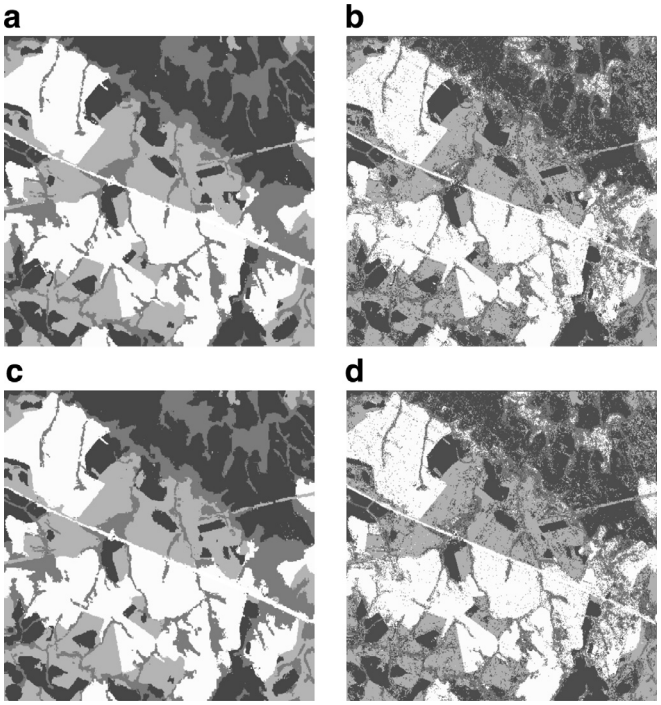

Fig. 4. CBERS-2B output images considering a training set with 20% of the entire image using: (a) Bayes-IA-CL, (b) Bayes-SW, (c) OPF-IA-CL and (d) OPF-SW.

Figs. 4–7 show the images classified by SW and IA-CL using a training set with 20% of the entire image considering CBERS-2B, Geoeye, Ikonos-2 MS and Landsat 5 TM satellites, respectively. We compared IA-CL against with SW only, since the latter one obtained the best results among the compared techniques.

A better performance of the proposed approach can be observed in all situations for both classifiers. Moreover, considering CBERS-2B, Geoeye and Ikonos-2 MS images, IA-CL obtained better results than all techniques using a smaller training set. Taking into account the images classified by OPF and Bayes, we can observe IA-CL was able to avoid misclassification in homogeneous regions of the image. Therefore, the images classified by SW look like noisy, although they are still better than naïve classification. Notice IA-CL

[3] In this work, we are using a 5 × 5-neighborhood system for SSL, MSSL-MR and MSSL-PY, and a 7 × 7-neighborhood for SW. We also employed 7 scales of decomposition for MSSL-MR, and 5 scales of decomposition for MSSL-PY. In regard to IA-CL, we used $\alpha = 2$ with the very same sampling process described in the Section 4.

[4] We employed an accuracy measure proposed by Papa et al. [17] that considers unbalanced datasets, which is often faced in land-cover image classification problems.

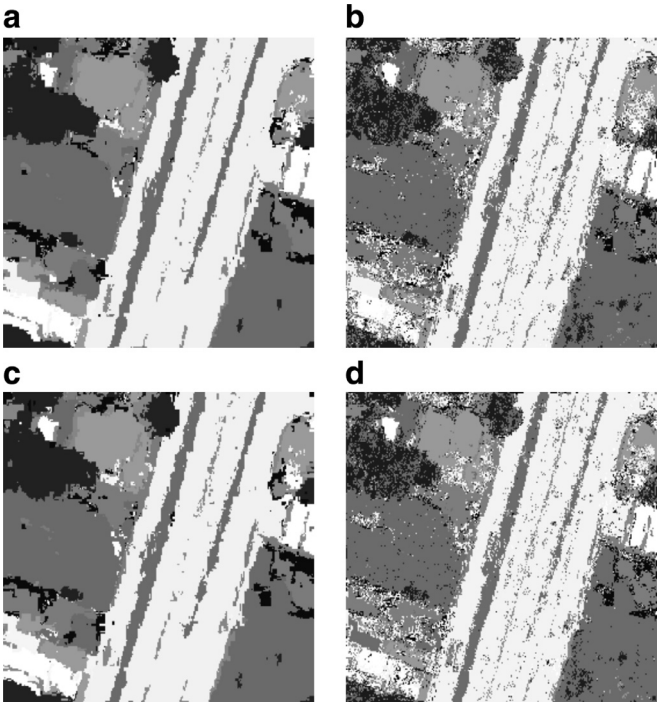

Fig. 5. Geoeye output images considering a training set with 20% of the entire image using: (a) Bayes-IA-CL, (b) Bayes-SW, (c) OPF-IA-CL and (d) OPF-SW.

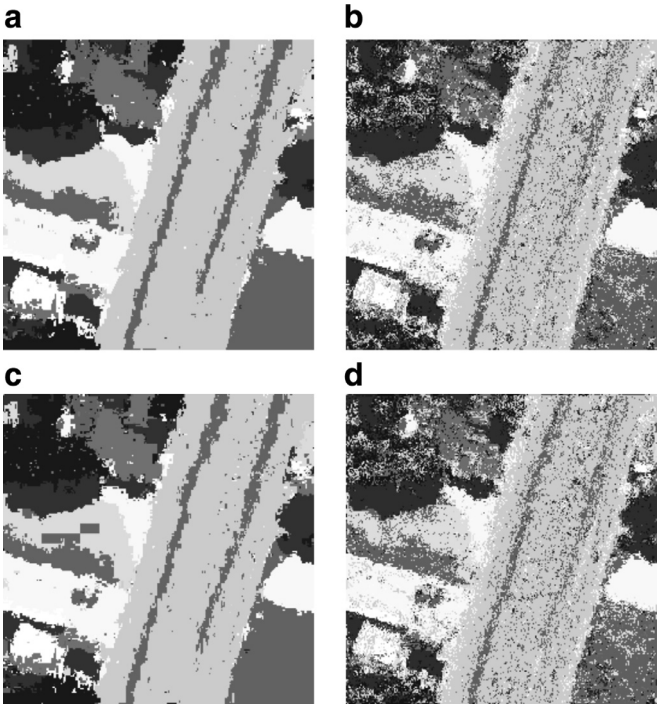

Fig. 6. Ikonos-2 MS output images considering a training set with 20% of the entire image using: (a) Bayes-IA-CL, (b) Bayes-SW, (c) OPF-IA-CL and (d) OPF-SW.

is more prone to errors in high-frequency regions, which is shortcoming often faced by the pattern recognition community.

Another interesting point concerns with the dimensionality of the feature space used to represent each pixel. Although SW requires $7 \times 7 \times 3 = 147$ features (we have 3 channels and a window of size 7×7), IA-CL needs $3 \times 7 \times 3 = 63$ features for CBERS-2B and Landsat 5 TM images only (we have 3 channels and

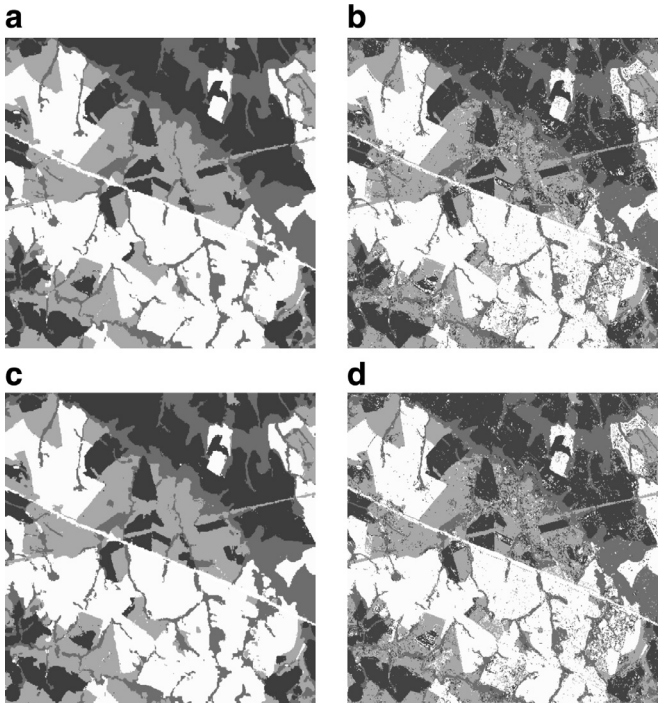

Fig. 7. Landsat 5 TM output images considering a training set with 20% of the entire image using: (a) Bayes-IA-CL, (b) Bayes-SW, (c) OPF-IA-CL and (d) OPF-SW.

3 features for each one of the 7 scales employed in this work), and $3 \times 6 \times 3 = 54$ features for Geoeye and Ikonos-2 MS images[5]. Additionally, as the number of scales is computed using a closed equation (Section 4), we do not have the problem of estimating such parameter.

7. Conclusions

The problem of contextual classification has been paramount in the last years, since the number of applications that are temporal- and spatial-dependent has increased in the last years. Although the techniques proposed to address this problem have obtained promising results, some of them require a high-dimensional feature space, which might be prohibitive for some parameter-dependent classifiers.

In this work, we presented a new contextual-based pattern classification technique that make use of the Interval Arithmetic tools in order to provide lower-dimensional feature spaces in a pyramid-based multi-resolution approach. The proposed approach was compared against with four sequential learning-based techniques using two distinct classifiers and three scenarios, obtaining the best results in all situations. The robustness of the proposed approach was evaluated in the context of supervised land-use classification employing images obtained from four satellites covering a Brazilian region. The visual results also showed the quality of the images classified by the proposed approach.

Acknowledgments

The authors are grateful to FAPESP grants #2014/16250-9 and #2015/50319-9, as well as CNPq grants #470571/2013-6, #306166/2014-3 and #487032/2012-8.

[5] We employed more scales for CBERS-2B and Landsat 5 TM images, since they have a bigger resolution.

References

[1] G. Alefeld, G. Mayer, Interval analysis: theory and applications, J. Comput. Appl. Math. 121 (1–2) (2000) 421–464.

[2] W.W. Cohen, V.R. Carvalho, Stacked sequential learning, in: Proceedings of the 19th International Joint Conference on Artificial intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005, pp. 671–676.

[3] T.G. Dietterich, Machine learning for sequential data: A review, in: T. Caelli, A. Amin, R.P.W. Duin, D. de Ridder, M. Kamel (Eds.), Structural, Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science, 2396, Springer Berlin Heidelberg, 2002, pp. 15–30.

[4] G.P. Drago, S. Ridella, Interval arithmetic multilayer perceptron as possibility-necessity pattern classifier, in: M. Marinaro, R. Tagliaferri (Eds.), Neural Nets WIRN Vietri-99, Perspectives in Neural Computing, Springer London, 1999, pp. 107–112.

[5] G.P. Drago, S. Ridella, Possibility and necessity pattern classification using an interval arithmetic perceptron, Neural Comput. Appl. 8 (1) (1999) 40–52.

[6] D. Dubois, E. Kerre, R. Mesiar, H. Prade, Fuzzy interval analysis, Fundam. Fuzzy Sets 7 (2000) 483–581.

[7] M. Fauvel, J. Chanussot, J. Benediktsson, A spatial–spectral kernel-based approach for the classification of remote-sensing images, Pattern Recognit. 45 (1) (2012) 381–392.

[8] C. Gatta, E. Puertas, O. Pujol, Multi-scale stacked sequential learning, Pattern Recognit. 44 (10–11) (2011) 2414–2426. Semi-Supervised Learning for Visual Content Analysis and Understanding.

[9] J. Kittler, J. Föglein, Contextual classification of multispectral pixel data, Image Vis. Comput. 2 (1) (1984) 13–29.

[10] R. Moore, W. Lodwick, Interval analysis and fuzzy set theory, Fuzzy Sets Syst. 135 (1) (2003) 5–9.

[11] R.E. Moore, Interval Analysis, Prentice-Hall, Englewood Cliff, New Jersey, 1966.

[12] R. E. Moore, Methods and applications of interval analysis, SIAM Studies in Applied Mathematics (1979).

[13] H.T. Nguyen, A note on the extension principle for fuzzy sets, J. Math. Anal. Appl. 64 (2) (1978) 369–380, doi:10.1016/0022-247X(78)90045-8.

[14] D. Osaku, R.Y.M. Nakamura, L.A.M. Pereira, R.J. Pisani, A.L.M. Levada, F.A.M. Cappabianco, A. Falcão, J.P. Papa, Improving land cover classification through contextual-based optimum-path forest, Inf. Sci. 324 (2015) 60–87.

[15] S.K. Pal, Fuzzy sets in image processing and recognition, in: Fuzzy Systems, 1992., IEEE International Conference on, 1992, pp. 119–126.

[16] J.P. Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, Efficient supervised optimum-path forest classification for large datasets, Pattern Recognit. 45 (1) (2012) 512–520.

[17] J.P. Papa, A.X. Falcão, C.T.N. Suzuki, Supervised pattern classification based on optimum-path forest, Int. J. Imaging Syst. Technol. 19 (2) (2009) 120–131.

[18] D.R. Pereira, R.Y.M. Nakamura, R.J. Pisani, J.P. Papa, Land-cover classification through sequential learning-based optimum-path forest, in: IEEE International Geoscience and Remote Sensing Symposium, IEEE Press, 2016, pp. 76–79.

[19] E. Puertas, S. Escalera, O. Pujol, Multi-class multi-scale stacked sequential learning, in: C. Sansone, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems, Lecture Notes in Computer Science, 6713, Springer Berlin Heidelberg, 2011, pp. 197–206.

[20] D. Ryabko, Pattern recognition for conditionally independent data, J. Mach. Learn. Res. 7 (2006) 645–664.

[21] F. Sampedro, S. Escalera, A. Puig, Iterative multi-class multi-scale stacked sequential learning: definition and application to medical volume segmentation, Pattern Recognit. Lett. 46 (0) (2014) 1–10.

[22] Y. Tarabalka, M. Fauvel, J. Chanussot, J. Benediktsson, SVM- and MRF-based method for accurate classification of hyperspectral images, IEEE Geosci. Remote Sens. Lett. 7 (4) (2010) 736–740.

[23] A. Wehmann, D. Liu, A spatial–temporal contextual markovian kernel method for multi-temporal land cover mapping, {ISPRS} J. Photogramm. Remote Sens. 107 (2015) 77–79.

[24] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (1945) 80–83.