ARTICLE IN PRESS
JID: PATREC [m5G;May 2, 2016;13:7]

Pattern Recognition Letters 000 (2016) 1–7

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 xxx

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

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.:

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

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