Pattern Recognition Letters 46 (2014) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Iterative multi-class multi-scale stacked sequential learning: Definition and application to medical volume segmentation ☆

Frederic Sampedro [a,*], Sergio Escalera [b,c], Anna Puig [b]

[a] Faculty of Medicine, Autonomous University of Barcelona, Barcelona 08193, Spain
[b] Faculty of Mathematics, University of Barcelona, 09007 Barcelona, Spain
[c] Computer Vision Center, Autonomous University of Barcelona, 08193, Spain

ARTICLE INFO

*Article history:*
Received 11 March 2014
Available online 21 May 2014

*Keywords:*
Machine learning
Sequential learning
Multi-class problems
Contextual learning
Medical volume segmentation

ABSTRACT

In this work we present the iterative multi-class multi-scale stacked sequential learning framework (IMMSSL), a novel learning scheme that is particularly suited for medical volume segmentation applications. This model exploits the inherent voxel contextual information of the structures of interest in order to improve its segmentation performance results. Without any feature set or learning algorithm prior assumption, the proposed scheme directly seeks to learn the contextual properties of a region from the predicted classifications of previous classifiers within an iterative scheme. Performance results regarding segmentation accuracy in three two-class and multi-class medical volume datasets show a significant improvement with respect to state of the art alternatives. Due to its easiness of implementation and its independence of feature space and learning algorithm, the presented machine learning framework could be taken into consideration as a first choice in complex volume segmentation scenarios.



## 1. Introduction and related work

Machine learning schemes have been widely applied in many image and volume segmentation scenarios [1,2]. A common property of image and volume data is the high context dependence condition of any pixel or voxel value, which breaks the common assumptions of most machine learning strategies regarding statistical independence of the entries. Although this fact could be ignored during the machine learning framework design, a number of works have tried to include it, showing a significant improvement on the global classification results.

Several strategies have been proposed for taking into account the contextual information during the learning process. When a particular set of prior assumptions about the context distribution can be defined, a number of learning schemes have been introduced, including graphical models [3–6], super-pixel methods [7–9], contextual priming [10] or Graph cuts and graph-based learning and spatial optimization [11,12]. As an example in 3D medical volume segmentation, in [13], the authors define a hierarchical structure based on Markov Random Fields. Also, in [14], the authors present a segmentation approach of the femoral head and proximal acetabulum from three dimensional (3D) CT data. This method is based on using boundary information within a Bayesian framework in a very specific medical imaging segmentation problem. In [15] a Bayesian framework is also used to model level sets for segmenting medical images.

However, most of these previous approaches either require a pre-established prior distribution of the spatial properties of the structures to be segmented, or consider a set of homogeneous features of the neighboring voxels. However, in many cases, such as the medical volume segmentation problem, the target regions to be segmented cannot be defined to appear spatially distributed within any consistent distribution nor to hold homogeneous properties. Given this fact, in this work we extract a set of standard features from medical volumes, and define a contextual learning approach based on discriminative classifiers, which will be able to learn those relevant dependencies and spatial relations within the training set. For this task, we focus on the stacked learning framework [16], and in particular, in the multiscale stacked sequential learning (MSSL) alternative presented in [17,18]. In those works, the authors extend the stacked sequential learning to include label predictions in the neighborhood of image pixels at different scales in the feature vector of 2D training samples, outperforming state-of-the-art approaches for 2D image segmentation. An advantageous property of the MSSL framework is its independence to any feature space or considered classifier.

☆ This paper has been recommended for acceptance by C. Luengo.

* Corresponding author. Tel.: +34 699805231.
*E-mail address:* fredsampedro@gmail.com (F. Sampedro).

http://dx.doi.org/10.1016/j.patrec.2014.05.002
0167-8655/

In this work, we first present a novel contextual definition of stacked learning using spatial 3D grids to compute feature vectors and label spatial relations between voxels. The feature vectors are trained using an ensemble of binary classifiers and combined within the error-correcting output codes framework (ECOC). On top of that, we build an iterative framework: iterative multiclass multiscale stacked sequential learning (IMMSSL). In each iteration within this iterative scheme, we treat the MSSL procedure as a black box where its output builds a new instance of the procedure in the iterative process. The refinement continues until the accuracy converges or an application-specific condition is met. Once this new learning scheme is defined, we present its performance results in three medical volume segmentation problems, which include two-class and multiclass segmentation scenarios. Results show that the proposed framework achieves encouraging performance results with respect to previous approaches in this scientific context.

The rest of the paper is organized as follows. Section 2 introduces the mathematical learning framework of the proposed IMMSSL scheme. Section 3 present the obtained performance results on three two-class and multiclass medical volume segmentation problems and Section 4 concludes the paper and points out some future directions.

## 2. The IMMSSL framework

In this section, we present the iterative multiclass multiscale stacked sequential learning scheme for volume segmentation. First, we review the stacked sequential learning (SSL) [16] and the multiclass multiscale stacked sequential learning (MMSSL) [17,18] schemes. Then, an improved iterative framework built on top of them in the context of 3D segmentation applications is presented (IMMSL).

### 2.1. SSL and MSSL frameworks

Given a two-class learning scenario with a training matrix ($X$) and its ground truth label vector for each of the entries ($g$), any classifier model $C_0$ can be trained using any supervised learning algorithm $h$:

$$C_0 = h(X, g)$$

Then, this classifier can be tested on the dataset, obtaining for each entry a binary classification predicted label $y$. Considering the fact that if a pixel or voxel has been classified with a particular label, its neighborhood points are more likely to be classified as such than the others. The approach taken by the SSL and MSSL models is to augment the original feature set adding new features related with the contextual information provided by $y$, and then train a new classifier model $C$ that would be able to learn the spatial context properties of the data and outperform previous segmentation results obtained by $C_0$.

In particular, the original feature set is extended with a set of contextual features ($Z$), computed from the base classifier ($C_0$) predictions ($y$) within a set of context scales ($s$) using a context function $J$. This process generates an extended training matrix $X'$. Note that when the set s of neighborhood scales is fixed to a single context lattice, the framework is referred as stacked sequential learning [16], and when $s$ includes several context lattice dimensions, it is referred as multiscale stacked sequential learning [17]. Finally, the contextual classifier $C$ will be trained using the extended training matrix $X'$. This extended feature set is defined as consisting of the original data feature set plus the set of features extracted from the contextual information provided by $J$, which in case of a binary classification problem its size is twice the cardinality of $s$, since for each context scale the percentage of predicted labels for each of the two possible class labels are computed [17]. This approach is summarized in Fig. 1.

Given that, the MSSL approach has shown better performance results than SSL [17], it has been used as the theoretical root for the rest of this work. Note the complete independence of any feature set, learning algorithm, or spatial distribution of the regions of interest within the SSL and MSSL frameworks. This property makes it especially appropriate to use in complex medical volume segmentation problems, since any medical imaging modality (CT, MRI, PET, US, etc.) differs from the rest in terms of spatial distributions, context properties and types of regions of interest.

### 2.2. Multi-class MSSL using error-correcting output codes

When dealing with a multiclass problem, we are interested to segment different regions within an image or volume. Given that most state-of-the-art learning strategies are defined to deal with two-class problems, extension to multi-class uses to be defined as an ensemble of binary classifiers. In this sense, the Error Correcting Output Code (ECOC) strategy has shown to be a powerful framework for the combination of classifiers to deal with multi-class data [19,20]. In short, given an $L$-class problem, a coding matrix $M \in \{-1, 0, +1\}^{L x n}$ is designed, where each column represents a binary classifier ($h_i$), the $L$ rows are defined as the codewords codifying each class $c_i$, $i \in [1, \ldots, L]$, and +1, $-1$ identifies the class membership for a binary classifier, being 0 if the class is not considered by the classifier. Then, for any given test instance $x$, its codeword $w$ is computed applying all binary classifiers, and its classification prediction is defined by the class with codeword at minimum distance given a distance metric (for instance the hamming distance [20]). An example in medical volume segmentation using 4 labels and one-vs-one ECOC design is shown in Fig. 2.

In the multiclass multiscale stacked sequential learning (MMSSL) [18], once the multiclass classification scheme has been defined (e.g. ECOC), it is included as a contiguous module after the binary classification stages, as shown in Fig. 3. Note the distinction between the first or base classification level ($C_0$) and the contextual level ($C$) that learns a new classifier from the extended matrix $X'$ containing multiple prediction labels.

### 2.3. Iterative MMSSL framework in volume segmentation scenarios

Our model proposal relies on the fact that if a contextual classifier can be trained from the information obtained from a base classifier, a new contextual classifier could also be trained using the actual predictions of this contextual classifier. This process can then be repeated iteratively. In this way, a larger set of complex spatial relations could be learned through the iterative refinement of each contextual classifier, and in consequence increase the generalization power of the whole system. Following this reasoning, the iterative multiclass multiscale stacked sequential learning is proposed as shown in Fig. 4.

The main idea is to progressively learn any contextual rules from the hits and misses of previous MMSSL classifiers. Thus, at

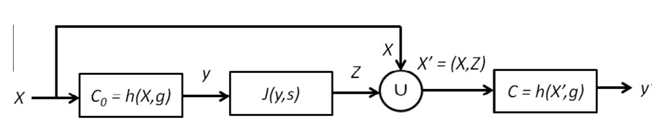


**Fig. 1.** SSL–MSSL approach in building a contextual classifier ($C$) extending the original feature set of $X$ using the predictions $y$ obtained from a base classifier $C_o$ and a contextual information extraction function $J$ in a set of neighbor scales $s$. With the extended training matrix $Z$, the contextual classifier outputs the final predicted labels $y'$.