Neurocomputing 152 (2015) 45–57

Contents lists available at ScienceDirect

Neurocomputing

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

Incremental classification of objects in scenes: Application to the delineation of images

Guillaume Bernard [a,b], Michel Verleysen [a], John A. Lee [a,b,*,1]

[a] Machine Learning Group – ICTEAM, Université catholique de Louvain, Belgium
[b] Molecular Imaging, Radiotherapy, and Oncology – IREC, Université catholique de Louvain, Belgium

ARTICLE INFO

Article history:
Received 15 April 2013
Received in revised form
16 October 2014
Accepted 5 November 2014
Communicated by Hantao Liu
Available online 14 November 2014

Keywords:
Multiclass classification
Incremental classification
Object recognition
Image segmentation

ABSTRACT

Usual multi-class classification techniques often rely on the availability of all relevant features. In practice, however, this requirement restricts the type of features that can be considered. Features whose value depends on some partial, intermediate classification results, can convey precious information but their nature hinders their use. A typical example is the identification of objects in a scene, where the distance from some yet unclassified object to some other that would already be identified earlier in the process. This paper proposes a generic method that solves classification problems involving such features in an incremental way. It proceeds by decomposing the multi-class problem into a sequence of simpler binary problems. Once a binary classifier gives an object its class tag, all features depending on this object are computed and appended to the list of known features. Experiments with both synthetic and real data, comprised of tomographic images, show that the proposed method is effective.

1. Introduction

Object recognition in images is a long-standing problem in the computer vision literature. It entails both segmentation and classification aspects. The proposed methods greatly vary, depending on the problem at hand and the data specificities. For instance, one way to segment an object consists in identifying its pieces such as proposed in [1]. In particular, the authors detect bodies in natural images by locating arms, heads, and legs. The method is effective but it aims at detection rather than actual segmentation (it yields a rectangular window encompassing the sought object). Several other methods provide tighter object contours and exploit image properties and features [2–4]. These methods achieve both segmentation and recognition (they label the image segments). Most of them are intended to work with natural images. In [2], a region-based energy functional is defined by individual segment potentials and inter-segment potentials. A two-steps hill climbing procedure minimizes the energy. The first step consists in giving a label to a group of pixel. The second step optimizes the region shape and updates its properties. The two alternate steps are repeated until convergence to a local minima of the energy. In [3], the authors proposed to extract a bag of segments. Those segments are extracted at different scales and locations. Non-overlapping segments are used to build tilings of the image (graphs that connect adjacent segments). The segmentation and labels of a new image are based on parameters learnt from previously labeled images. The most probable tiling is selected and its associated labels are then copied and attached to the tiles of the new image. In [5], the tiling of different images from the same scene is used to make a co-labelling of the image. All the images are jointly annotated, thereby giving more consistent values. In computer vision, labels are often given to (groups of) pixels but sometimes the whole image can also be labeled. This approach is investigated in [4]. Richer label information is expected to improve the results but such a method has higher requirements for the data collection process. Another method to improve the annotation of an image is label propagation, where the obtained labels are corrected by propagation. An interesting use of label propagation can also be seen in [6]. To be able to identify several drosophila embryo stages on one image, the authors use the shape of the embryos to get a first label. Then, they use a similarity measure based on patterns in the embryos to make the label propagation.

This paper follows a different approach, which is intended to solve a very specific problem. The main assumption is that data consists of similar scenes, all including the very same set of objects. Medical image segmentation, for example, enters within this framework: all patients are imaged following mostly the same protocol, share the same anatomy, but differ in their size, weight, and morphology. In order to interpret medical images, physicians

* Corresponding author.
E-mail addresses: guillaume.bernard@uclouvain.be (G. Bernard), michel.verleysen@uclouvain.be (M. Verleysen), john.lee@uclouvain.be (J.A. Lee).
[1] The Author is a Research Associate with the Belgian fund of scientific research F.R.S.-FNRS.

http://dx.doi.org/10.1016/j.neucom.2014.11.020
0925-2312/

proceed step by step. They typically start by using the little available information to label a few first organs. By doing so, they can deduce new pieces of information, which were initially unavailable, allowing them to recognize new organs, and so forth.

This paper formalizes such an incremental classification process and provides two slightly different methods of solving it, both inspired by a ‘divide and conquer’ approach. They consist in solving a succession of usual, binary classification problems, which are fed with the available features at the time of their respective execution. The two proposed methods actually differ in the way they sequence the ordinary classifiers. As a proof of concept, the two methods are used to solve object recognition problems involving synthetic and real tomographic images.

The remainder of this paper is organized as follows. Section 2 formalizes the problem of incremental classification. It also defines the various terms and symbols used throughout this paper. Section 3 describes the proposed method of decomposing incremental multi-class problems into a sequence of simpler binary problems. In particular, it details two different ways to determine the sequence of these subproblems. Section 4 presents the experiments and their results. Eventually, Section 5 draws the conclusions and sketches some perspectives for future work.

2. Incremental classification: formalization of the problem

Classification is the task of labeling objects or data items, according to known features, and based on previously seen examples. A typical classification problem includes a learning set (examples of data items for which the class label is known) and a query (a set of instances for which the label has to be attributed). Both the learning set and the query consist of feature vectors, that are supposed to be drawn from the same distribution, so that the learning set is representative of the query. All features are supposed to be known at the time of resolving the query. If this assumption has the merit to frame the classification problem, it can be constraining in practice. As an example, let us consider our visual system, when it analyses a scene. Our brain is able to recognize and tag objects of the scene, but not all features of the objects are known from the beginning. In particular, our eyes use complex features such as the spatial or geometrical relationships between the objects in the scene. Such features are not known if one of the objects taking part in the relationship has not been labeled yet. Therefore, solving the whole problem requires not only information collection, but also some reasoning: simpler subproblems must be solved in an incremental way to progressively build the missing pieces of information.

For instance, a child who does a jigsaw puzzle solves such a problem. Another example is a physician who looks at a radiographic or tomographic image. The diagnostic depends on the interpretation of the image and therefore on the sequential recognition of the depicted organs.

All these problems share several characteristics. Most of them are visual problems, in which objects must be recognized or differentiated. Sometimes the objects are intrinsically dissimilar, making the solution obvious. Sometimes the objects bear some confounding similarity and differentiating them requires additional, extrinsic information coming from their environment and their relationship with other objects.

In order to formalize the problem, a few terms are defined hereafter.

Scene ($\mathcal{S}$): A scene is a picture of objects living in a N-dimensional space. The scene can be encoded in various ways. Here we assume that an image of the scene is available, either as a projection (like a 2D picture taken by a digital still camera) or a full 3D image (like tomographic images in medical imaging). As an example, let us consider a scene composed of three objects: a green bike, a red ball, and the ground (see Fig. 1).

Objects (*set* $\mathcal{O}$ *composed of* o_j): The objects are the main elements of the scene. All the objects of the scene must be identified at the end of the iterative process. In the example, they are the bike, the ball, and the ground. In practice, to determine the border of the object we use segmentation algorithms [7–10]. Due to noise of varying illumination, those algorithms tend to over-segment the image. The object is therefore fragmented into several pieces.

Pieces (*set* $\mathcal{P}$ *composed of* p_i): The pieces are the over-segmented parts of the objects. $\mathcal{P}$ includes all pieces of the segmented image of the scene. In the example, the bike can be split in several pieces: two wheels, a frame, a handlebar, a saddle, etc. The ball is quite uniform so it would only be composed of one piece. The ground can be split in two pieces: a piece is illuminated by sunlight while the other is shadowed.

O-features and P-features: Pieces and objects have several features allowing their identification. In the example, the pieces have a color: the frame of the bike is green, the ball is red, the saddle and the ground in the shadow are both gray. The size of the objects (and pieces) can also be used as features. For this paper, $\mathcal{O}$-features refer to the features of an object and $\mathcal{P}$-features refers to the features of a piece. The features can be of any type: Boolean, real, categorical, etc. The objects, pieces, and their respective features are represented in Fig. 2.

Those elements and features do not suffice to describe a scene. The relationships between pieces and objects must be defined as well.

Labels ($\mathcal{L}$: *set of* $\mathcal{L}$-*edges*): The labels give the membership of a piece to an object. In the example, the saddle belongs to

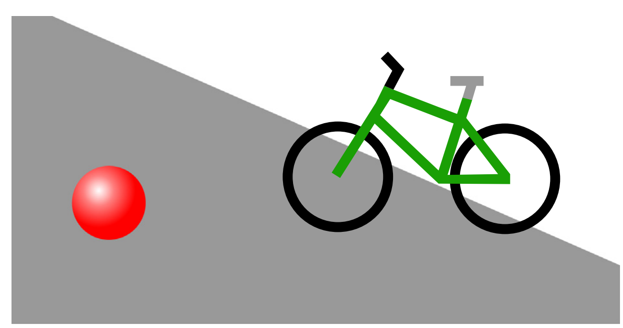

Fig. 1. Example of scene, composed of a red ball, a green bike, and partially shadowed ground. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

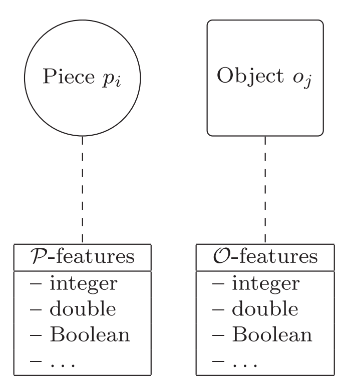

Fig. 2. Pieces and objects are the two main elements of the scene. They are represented with there respective features.