Information Sciences 340–341 (2016) 17–26

Contents lists available at ScienceDirect

Information Sciences

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

An effective learning strategy for cascaded object detection

A. Bria, C. Marrocco, M. Molinara, F. Tortorella*

Department of Electrical and Information Engineering, University of Cassino and L.M., Via Di Biasio 43, Cassino (FR) 03043, Italy

ARTICLE INFO

Article history:
Received 28 January 2015
Revised 7 August 2015
Accepted 4 January 2016
Available online 11 January 2016

Keywords:
Pattern recognition
Image processing
Object detection
Cascade
Unbalanced data
Ranking

ABSTRACT

To distinguish objects from non-objects in images under computational constraints, a suitable solution is to employ a cascade detector that consists of a sequence of node classifiers with increasing discriminative power. However, among the millions of image patches generated from an input image, only very few contain the searched object. When trained on these highly unbalanced data sets, the node classifiers tend to have poor performance on the minority class. Thus, we propose a learning strategy aimed at maximizing the node classifiers ranking capability rather than their accuracy. We also provide an efficient implementation yielding the same time complexity of the original Viola–Jones cascade training. Experimental results on highly unbalanced real problems show that our approach is both efficient and effective when compared to other node training strategies for skewed classes.

1. Introduction

Detecting objects in images and videos is a crucial task in many real-world problems, ranging from face detection in images to pedestrian detection on roads, from biomedical image analysis to video surveillance applications. From the pattern recognition point of view, object detection can be cast as a classification problem in which one has to distinguish between the object class and the "non-object" class. In the last years, most of the object detection approaches proposed in the literature rely on the concept of "objectness" [1], for which objects in an image are characterized by a well-defined boundary, a different appearance from the surroundings and sometimes by uniqueness [1]. This is particularly true for objects that can be described by a structure made of more connected elementary parts [11], where each part can capture and describe different properties of the object. This concept, however, cannot be applied to all the categories of "things" present in the images. Several real-world applications [2,7,26,29,30] deal with objects that are not distinguishable from the background since they are not neatly different from their surroundings and are not unique within the image. This situation can depend on many factors such as the characteristics of the employed sensor, the size of the objects or the resolution of the images at hand. Moreover, every real-world detection task requires processing a huge number of pixels and this can involve a computational load not easy to bear specially when the classifier must be designed under computational constraints [23].

In this regard, a commonly adopted solution, originally proposed by Viola and Jones [32], is to employ an ensemble of classifiers structured as a *cascade* of dichotomizers with increasing complexity. Such an approach allows each dichotomizer in the cascade to deal with only a part of the non-object class, thus parting the complexity of the whole problem among the classifiers. Specifically, the first stages of the cascade are built to reject the most distinguishable background regions, while the last stages are specialized to discriminate between actual object and the most confusing background patches (see

* Corresponding author. Tel.: +39 07762993605.
E-mail addresses: a.bria@unicas.it (A. Bria), c.marrocco@unicas.it (C. Marrocco), m.molinara@unicas.it (M. Molinara), tortorella@unicas.it, fr.tortorella@gmail.com (F. Tortorella).

http://dx.doi.org/10.1016/j.ins.2016.01.021

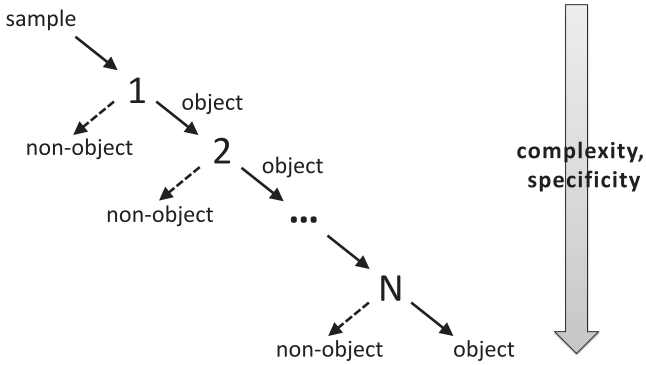

Fig. 1. Scheme of the cascade structure with N nodes.

Fig. 1). This aims at reducing the number of false positives produced by the detector and concentrate the computational complexity of the system on the last classifiers of the cascade.

Moreover, another problem are the skewed class priors since the overwhelming majority of sub-windows in an image do not contain the target object. In the original Viola–Jones approach, the cascade dichotomizers are trained with *AdaBoost* [13], which is not designed to face the asymmetry between the classes. In fact, *AdaBoost* aims at maximizing the dichotomizer accuracy, a quantity related to classification error, that is significantly biased by skewed class priors. Indeed, accuracy is not a correct measure for handling rare cases that have less impact than common cases [38]. As a consequence, learning based on accuracy may lead to poor minority-class performance [37].

A number of approaches have been proposed to alleviate this problem, including reformulations of *AdaBoost* that take into account classification error costs [33,35] or artificially rebalance class distributions (such as over-sampling or under-sampling of training set) [6,25], and methods that modify the structure of the cascade at the cost of a higher computational complexity [19,36,39].

Here we present an approach to use rank metrics instead of accuracy metrics for training the dichotomizers in the cascade. Rank metrics have been successfully used in the information retrieval domain [15,17,42–44] to learn ranking models that rank new items similarly to rankings in the training data. Specifically, rank metrics are more concerned with the relative ordering of cases than with making absolute predictions for cases and thus place more emphasis on learning to distinguish classes than on learning the internal structure of classes [4].

To this end, we propose a training strategy for the dichotomizers based on a reformulation of *RankBoost* for bipartite ranking problems [12], suitably modified to be embedded in a cascade structure. This solution allows the detector to take advantage of the benefits of the standard cascade framework and to effectively face the class asymmetry.

This paper is a considerably extended version of our conference report [3]. We here present the full details of our method, an efficient implementation, and an extensive evaluation on four real-world applications from the fields of computer vision and biomedical image analysis: (i) face detection; (ii) pedestrian detection; (iii) detection of microcalcifications on digital mammograms; and (iv) detection of microaneurysms in digital fundus images. All the examined applications are characterized by a considerable class imbalance and the obtained results show that the proposed approach provides performance similar or better than the compared methods.

We start with a brief review of related work to face class asymmetry in the cascade dichotomizers (Section 2) and then present the underlying concepts (Section 3), an efficient implementation (Section 4), and experimental evaluation (Section 5) of our method, followed by a summary of the conclusions that can be derived from the results (Section 6).

2. Related work

The approaches that have been proposed to face class imbalance in the cascade dichotomizers can be attributed to two different groups: (i) cost-based strategies that introduce classification error costs; (ii) sampling-based strategies that attempt to balance the class distributions by undersampling the majority class or by oversampling the minority class.

2.1. Cost-based strategies

This category includes the methods that are based on reformulation of *AdaBoost* with a cost assigned to false negatives greater than to false positives. Let us briefly recall the *AdaBoost* learning algorithm. A weak learner $h_\tau(\cdot)$ is selected in each of a series of rounds $\tau = 1, 2, \ldots,$ so as to minimize the *weighted exponential loss* $\sum_j D_\tau(j)\exp\left(-y_j h_\tau\left(\mathbf{x}_j\right)\right)$, where $D_\tau(j)$ is the weight on the jth sample $\mathbf{x}_j$ and $y_j \in \{-1, 1\}$ is the class label of $\mathbf{x}_j$. It is easy to see that, in the case of highly skewed classes, such sum is dominated by the error produced on the majority class and thus the choice of the weak learner is not optimal for predicting the minority class. To alleviate this problem, Viola and Jones proposed *AsymBoost* [33], where each sample $\mathbf{x}_j$ is pre-weighted at each round with an asymmetric cost $\exp\left(\frac{1}{T} y_j \log\sqrt{k}\right)$ where T is the total number of rounds of