Information Fusion 28 (2016) 24–43

Contents lists available at ScienceDirect

Information Fusion

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

Diversity-aware classifier ensemble selection via f-score

Ingrid Visentini, Lauro Snidaro *, Gian Luca Foresti

Department of Mathematics and Computer Science, University of Udine, via delle Scienze 206, 33100 Udine, Italy

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form 14 July 2015
Accepted 15 July 2015
Available online 31 July 2015

Keywords:
F-score
Classifiers selection
Classifiers fusion
Tracking via classification
Online tracking

ABSTRACT

The primary effect of using a reduced number of classifiers is a reduction in the computational requirements during learning and classification time. In addition to this obvious result, research shows that the fusion of all available classifiers is not a guarantee of best performance but good results on the average. The much researched issue of whether it is more convenient to fuse or to select has become even more of interest in recent years with the development of the Online Boosting theory, where a limited set of classifiers is continuously updated as new inputs are observed and classifications performed. The concept of online classification has recently received significant interest in the computer vision community. Classifiers can be trained on the visual features of a target, casting the tracking problem into a binary classification one: distinguishing the target from the background.

Here we discuss how to optimize the performance of a classifier ensemble employed for target tracking in video sequences. In particular, we propose the F-score measure as a novel means to select the members of the ensemble in a dynamic fashion. For each frame, the ensemble is built as a subset of a larger pool of classifiers selecting its members according to their F-score. We observed an overall increase in classification accuracy and a general tendency in redundancy reduction among the members of an f-score optimized ensemble. We carried out our experiments both on benchmark binary datasets and standard video sequences.

1. Introduction

It is well known that the fusion of an ensemble of “weak” independent classifiers can lead to substantial performance improvements with respect to a single monolithic classifier [21]. The term “weak” is used to indicate a classifier that is not particularly specialized or trained for the problem at hand (i.e. it is sufficient that classification performance be slightly better than random guessing). These ensembles can be employed in a broad variety of applications, from medical imaging [48] to network security [16], from biometric person identification [24] to remote sensing [59], in a large range of real-world domains [40].

To fuse classifiers a large number of possible rules can be used [47]: for instance, sum and product [25], Bagging [5] and Boosting [14], Random Subspaces [22], or oracles [33]. Considering couples of classifiers, mutual information [44], Q statistic [60], diversity-based criteria [32,56] or correlation, for instance, can represent valid pairwise measures that consider their independence to merge their outputs.

To save computational time, an option is to employ only a selection of classifiers instead of the entire set [57]. The selection procedure is aimed at forming a reduced ensemble by choosing within a pool the subset of classifiers that maximizes the performance [30] or, alternatively, reduce the error. This approach is often applied to features [20] to decrease, for instance, the dimensionality of the input space or to choose a more robust subset, but it is also used for classifiers [1], to achieve better performance or to satisfy real-time constraints. In this context, a classifier combination strategy that links together selection and fusion includes switching between fusion and selection [30,50,12].

The recent development of online learning methods [39,43,34] has opened the possibility to build on-the-fly a classifier ensemble and to train it with incoming samples in an unsupervised manner and without any prior knowledge of data distribution. These techniques are based on a evolution of the original Boosting [55] algorithm and rely on a fixed size ensemble of classifiers, whose weights are continuously updated according to some statistical information on observed samples. However, for instance, the Online Boosting technique can present an optimistic view of the classifiers behavior, scoring only the distinction between correctly and wrongly labeled (classified) samples without considering the skewness of the training set (see [15] for a discussion on ensembles

* Corresponding author.
E-mail address: lauro.snidaro@uniud.it (L. Snidaro).

http://dx.doi.org/10.1016/j.inffus.2015.07.003
1566-2535/

for the class imbalance problem); assessing the performance of the classifiers in presence of an unbalanced number of training samples can be misleading.

For this reason, Pham and Cham [46] proposed an asymmetric online boosting algorithm, where both a parameter k, that takes into account the asymmetry of the class labels presented to the classifiers, and the number of false/true positives and false/true negatives are considered in the tuning of the coefficients of the linear combination of the classifiers. Even if the problem of unbalanced classes is handled, the application of the entire pool of classifiers can be still computationally expensive.

1.1. Online classification for video tracking

An interesting application of online learning methods is target tracking in video sequences, that recently has received a new boost thanks to the tracking via classification concept [8,18,45,53,38]. The idea is that classifiers can be trained on the visual features of a target, casting the tracking problem into a binary classification one: distinguishing the target from the background. In the vast majority of tracking applications, the target changes its appearance as it moves within the field of view of a video sensor due to rotations (of the target and/or camera) and perspective distortions. For this reason the model learned by the classifiers should be updated at every new frame in a continuous detect ↔ update cycle. The recent availability of methods for online training classifier ensembles on incoming data, like Online Boosting [18], has thus stoked the interest for this type of tracking instrument. The advantages over existing tracking methods are clear:

- the ensemble can be trained on heterogeneous features (e.g. color features, texture, motion, etc.) thus improving the robustness of the detector.
- being trained on a specific object, it works as a detector of the particular instance. In the case of multiple objects in the scene, each of them is tracked by a dedicated ensemble (i.e. trained on the target's features).

Recent works include Avidan's Adaboost-based tracker [3], that exploits features associated to every pixel. However, the work uses the classic Adaboost algorithm and does not learn online the appearance of the target. In [10] the most discriminative color features to separate the target from the background are chosen by applying a two-class variance ratio to log likelihood distributions computed from samples of object and background pixels. In a later work, heterogeneous features have been combined adopting the same fusion method [42]. In these two works the features are ranked and selected afresh for each frame without considering past history (i.e. how features performed in the previous frames).

In [18] the Online Boosting technique devised by Oza [39] is adapted for visual target tracking. Albeit this idea is effective, since it uses the online ensemble learning paradigm, it employs an architecture that relies on a fixed cardinality ensemble. No selection is applied and this can be detrimental for real-time constraints.

1.2. Algorithm outlook

In this work, we propose a new criterion based on the F-score measure to select classifiers from a set of constantly updated ensemble members (Fig. 1). This criterion has been used in [9] applied to SVM, but its application in online learning is still unexplored to the best of our knowledge.

1. INITIALIZATION: The full ensemble members are supervisedly trained with a set of labeled samples. This initialization is done once at the startup.

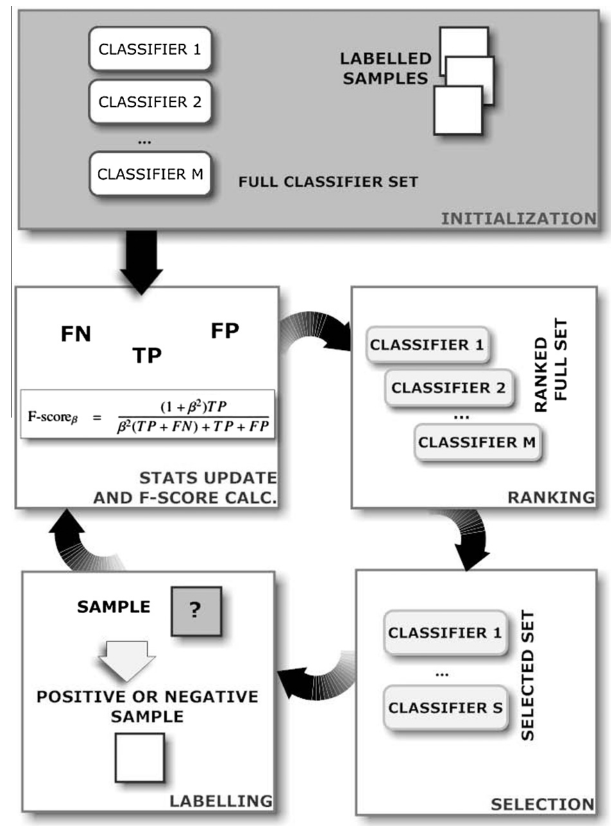

Fig. 1. Architecture of the proposed approach for selecting classifiers on-line, based on their F-score measure. The loop is described in detail in Section 1.2.

2. STATS UPDATE: The statistics (TN and TP, FN and FP, precision, recall and F-score) of each classifier of the full pool are individually updated. Since this step is a matter of storing a few variables the computation for this step is fast.
3. RANKING: The members are ranked in descending order using their revisited F-score value.
4. SELECTION: The classifiers for forming the reduced pool can be selected, as presented here in the paper, for being within the first S classifiers in the ranking.
5. LABELING: The selected ensemble classifies a new unlabeled testing/validation sample. The labeling of the sample is performed by the selected set only, while the other ensemble members are not considered in this phase.
6. LOOPING: While there are test samples available, repeat all the steps from (2).

This general proposed approach can be used in online and offline datasets. We will test it in both cases:

- In the (offline) case of UCI datasets, we train the ensemble members with a minimal training set of randomly picked samples (1/3 the size of the dataset). We then re-compute the F-score based ranking for each new validation sample (2/3 of the dataset). The validation samples are processed one-by-one.
- In the (online) case of a video sequence, where data is continuously streaming in, the ensemble is trained on a small amount of initial frames, where the positive samples are manually located as image patches (corresponding to the target) in a semi-supervised fashion. The validation samples are then found and labeled on-the-fly by the selection classifiers, which analyze the video stream frame-by-frame and picking the most likely image patch containing the target. In this case, the found