Pattern Recognition Letters 55 (2015) 35–41

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Structured multi-class feature selection with an application to face recognition[☆]

Luca Zini, Nicoletta Noceti[*], Giovanni Fusco, Francesca Odone

DIBRIS, Università di Genova, via Dodecaneso 35, Genova 16146, Italy

ARTICLE INFO

Article history:
Received 17 February 2014
Available online 22 July 2014

Keywords:
Multi-class structured feature selection
Group LASSO
Face recognition

ABSTRACT

In this paper we address the problem of structured feature selection in a multi-class classification setting. Our goal is to select groups of features meaningful to all classes simultaneously, and to this purpose we derive a new formulation of Group LASSO – the *MC-GrpLASSO* – and a solution of the obtained functional based on proximal methods. We then apply the algorithm to a typical multi-class problem – face recognition. On this respect we focus on finding an effective and fast to compute (that is, sparse) representation of faces, detected in low quality videos of unconstrained environments. We start from a classical over-complete representation based on Local Binary Patterns (LBPs), descriptors endowed with a characteristic internal structure that can be preserved by selecting features in groups. We present an extensive experimental analysis on two benchmark datasets, MOBO and Choke Point, and on a more complex set of data acquired in-house over a large temporal span. We compare our results with state-of-the-art approaches and show the superiority of our method in terms of both performances and sparseness of the obtained solution.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Many crucial aspects of image understanding can be formulated as classification problems which often involve more than two classes. In image classification a large body of research has been devoted to finding meaningful image features. Such features inherit a large extent of the spatial redundancy typical of images, and are often endowed with an internal structure (for instance due to the fact that groups of features need to be computed together because they refer to the same image portion or to the same scale). Since computer vision application often call for real-time performances, the richness of image representations need to be balanced with automatic methods for dimensionality reduction. Here feature selection plays an important role, as it allows us to learn what features are meaningful for a given problem, and to discard the remaining features from the run-time computation.

In this paper, we address the problem of finding a sparse representation which preserves the data internal structure, and propose a new formulation of the Group LASSO functional [33] – we call *MC-GrpLASSO* – to cope directly with a multi-class setting. The new formulation allows us to select groups of features that simultaneously discriminate among *all the classes.*

We then apply the algorithm to face recognition, learning a *very sparse face representation* from data. From the application standpoint, we are interested in unconstrained settings where people are free to move as they like. Videos are captured by low quality cameras, such as video-surveillance sensors or mobile devices. We start from an overcomplete description based on Local Binary Patterns (LBP) [34,1], which have been shown effective for analyzing textures and patterns in low quality images. LBPs are computed at different scales and aspect ratios, then are subject to a feature selection step with the goal of reducing their redundancy, while improving the computational efficiency of the method at run time. On top of these modules, we build an efficient face recognition pipeline, analyzing a face image and associating it one among *N* previously learnt identities.

Multi-class feature selection is still largely unexplored and few are the applications to the computer vision domain. A reference paper may be [5], while some applications can be found mostly from the computational biology field [5,37]. In computer vision the problem of feature selection is often addressed by defining a set of binary problems. In this case we may obtain representations which are meaningful for a given class. In the case of faces this approach is normally referred to as *identity-based* recognition [8,12] and, follows the intuition that different people may be

[☆] This paper has been recommended for acceptance by G. Sanniti di Baja.

[*] Corresponding author. Tel.: +39 010 3536626; fax: +39 010 3536699.
E-mail addresses: luca.zini@unige.it (L. Zini), nicoletta.noceti@unige.it (N. Noceti), giofusco@gmail.com (G. Fusco), francesca.odone@unige.it (G. Odone).

http://dx.doi.org/10.1016/j.patrec.2014.07.004
0167-8655/© 2014 Elsevier B.V. All rights reserved.

characterized by different types of features. Although this idea has a very convincing application ground, since it is true that some features may be more meaningful for one person than they are for another (cheeks may be meaningful for a person wearing a beard, the forehead for a subject with hair combed in a peculiar way, …), this approach does not favor system scalability as the number of classes grows, since it requires at run time the computation of a large number of different representations. Indeed, it is usually adopted mainly for face authentication. The approach we take in this work is somehow dual, since we try to find a compact description able to represent the majority of people and possibly to be extended to new people with a limited impact on the recognition performances.

In summary, the contribution of our work is thus twofold: (i) a new formulation of Group LASSO to directly model multi-class problems, and (ii) an efficient and modular face recognition pipeline. We discuss the recognition performances of our method on two benchmark datasets of low-quality videos – MOBO [11] and Choke Point [30] – and on the R309 dataset acquired in-house, of higher complexity. The experimental analysis shows that our method outperforms other state-of-art approaches on all the three datasets. We also report an analysis on the applicability of the method on a larger number of classes with the FERET dataset [20].

The remainder of the paper is organized as follows. Section 2 summarizes related works, Section 3 describes the multi-class feature selection method we propose, while Section 4 presents the algorithms and the software toolbox, which is available for download. Section 5 focuses on the entire face recognition pipeline. The experimental analysis on the different datasets is detailed on Section 6, while Section 7 is left to a final discussion.

2. Related works

Few works in the literature are, up to now, devoted to *multi-class feature selection*. We mention [5] which extends Recursive Feature Elimination (RFE) to the multi-class case, [26] based on probabilistic multi-class SVM, [21] where the authors propose feature selection methods based on the analysis of the kernel matrix and extend the methods to the multi-class case by applying an appropriate target kernel which maps a multi-class to a binary problem. A multi-class extension of Adaboost was included in the original paper [10] and has been recently considered in a grid-based implementation for learning face features [36]. This work mainly focuses on computational issues, as it is well known that Adaboost training phase may be computationally heavy, and it will be in particular in the multi-class case.

Among the applications of multi-class feature selection in the computer vision domain it is worth mentioning multi-class RFE applied to the classification of textures [6] — a problem recently addressed also by non-parametric statistics [22]. Multi-task transfer learning has been used to boost the performances of multiple object detection, by selecting features meaningful to all classes [28].

As for face recognition, in general feature selection is addressed as a combination of binary methods. Often feature selection is based on Adaboost: in [34] the authors start off with an overcomplete set of LBP histograms computed in different regions of face images. Then, they use Adaboost to select the most significant features. The work in [31], instead, introduces the Jensen–Shannon Boosting (JSBoost) algorithm, a modification of AdaBoost based on the Jensen–Shannon (JS) divergence, that is shown to provide a more appropriate measure of distance between examples of two different classes. In [27], the authors consider Gabor features to extend Adaboost by incorporating mutual information. They achieve a lower training error rate with respect to the standard implementation. The Gabor features are also adopted in [23], where Real Adaboost is introduced to perform the selection wavelets. A genetic based approach is instead proposed in [25], where the selection of the LBP regions is based on Genetic & Evolutionary Computing (GEC). Often, a *binary approach to the feature selection* problem is motivated by the interest for identity-based descriptions [8,12] tailored for the face authentication problem. Instead, we consider the problem of finding a representation which is compact and appropriate for a plurality of identities, in the literature referred to as universal descriptions. Universal description, common to all identities, give us an advantage in terms of scalability when new identities need to be enrolled in the system. To this purpose our method is specifically designed for multi-class settings.

Feature selection can be cast in the more general problem of finding appropriate image representations for an effective classification. Although the scope of this problem is rather large, if we focus on face recognition in particular, it is worth mentioning holistic approaches, and among them eigenfaces, see [29] and fisherfaces, see [3], component-based methods [15], local approaches as [8,12] and, more recently, adaptive approaches based on sparse coding [32]. Also the interested reader may refer to [35].

3. Regularized multi-class feature selection

Before describing the multi-class feature selection method we propose, we formulate the *feature selection problem*. Given a training set $(\mathbf{x}_i, y_i), i = 1, \dots, n$, with $\mathbf{x}_i \in X \subseteq R^m$ $y_i \in R$, and a dictionary $\mathcal{D} = (\phi_j)_{j=1}^p$ which is a collection of atoms or features, we consider a generalized linear formulation of the input–output relationship

$$\sum_{j=1}^{p} \phi_j(\mathbf{x}_i)\beta_j = y_i \tag{1}$$

or, in matrix form $\Phi\beta = \mathbf{y}$, where Φ is the feature matrix defined as $\Phi_{ij} = \phi_j(\mathbf{x}_i)$ of size $n \times p$, $\mathbf{y} \in \{-1, +1\}^{n\times 1}$ is the output vector, and $\beta \in \mathbb{R}^{p\times 1}$ is the vector of weights. The goal of feature selection is to find a *sparse* β which approximates well the input–output relationship.

Typically, in high dimensional domains, the dictionary matrix Φ is large and rectangular, with $p \gg n$. Thus, the linear system is underdetermined, thus algebraic solutions to the linear system are not feasible. Also, dictionaries are often over-complete [19] and the correlation between groups of features make the system ill-conditioned. A solution to the system can be found by introducing some form of regularization. A classical way is to adopt a Tikhonov regularization, formalized as follows

$$\beta^* = \arg\min_{\beta} \left(\|\mathbf{y} - \Phi\beta\|_2^2 + \mu\|\beta\|_2 \right), \tag{2}$$

where the first term is a *data fidelity term* while the second one is a *penalty term*.

LASSO and elastic net. The Tikhonov penalty term ensures a smooth solution β^* where all components will be different than zero. Instead, if a L_1 penalty is included in the functional, we obtain a sparse solution, which also provide us with a selection of the most meaningful features for the problem:

$$\beta^* = \arg\min_{\beta} \left(\|\mathbf{y} - \Phi\beta\|_2^2 + \mu\|\beta\|_2^2 + \tau\|\beta\|_1 \right). \tag{3}$$

The combination of the two penalties, known in the literature by the name of *elastic net* [38,7] combines the effects of obtaining a sparse and unique solution. The sparsity term, controlled by the parameter τ, measures the amount of features associated with a non zero coefficient, while the L_2 term restores the convexity of the functional and guarantees a unique solution. In this case groups

Download English Version:

https://daneshyari.com/en/article/536283

Download Persian Version:

https://daneshyari.com/article/536283

Daneshyari.com