Pattern Recognition 48 (2015) 2279–2289

Contents lists available at ScienceDirect

Pattern Recognition

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

Multi-label learning with missing labels for image annotation and facial action unit recognition

Baoyuan Wu [a], Siwei Lyu [b], Bao-Gang Hu [a], Qiang Ji [c,*]

[a] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China
[b] Department of Computer Science, University at Albany, SUNY, NY 12222, USA
[c] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form
29 November 2014
Accepted 26 January 2015
Available online 12 February 2015

Keywords:
Multi-label learning
Missing labels
Image annotation
Facial action unit recognition

ABSTRACT

Many problems in computer vision, such as image annotation, can be formulated as multi-label learning problems. It is typically assumed that the complete label assignment for each training image is available. However, this is often not the case in practice, as many training images may only be annotated with a partial set of labels, either due to the intensive effort to obtain the fully labeled training set or the intrinsic ambiguities among the classes. In this work, we propose a method for multi-label learning that explicitly handles missing labels. We train classifiers with the multi-label with missing labels (MLML) learning framework by enforcing the consistency between the predicted labels and the provided labels as well as the local smoothness among the label assignments. Experiments on three benchmark data sets in image annotation and one benchmark data set in facial action unit recognition demonstrate the improved performance of our method in comparison of several state-of-the-art methods.

1. Introduction

In multi-label learning, each example can be associated with many classes simultaneously. The typical examples include image annotation [1] and the recognition of facial action units (AUs) in facial images [2]. More formally, the complete labels of an example $\mathbf{x}$ over m classes, $\{c_1, \ldots, c_m\}$ are represented as a vector $\mathbf{z} \in \{-1, 1\}^m$, where a positive value indicates that the example belongs to the corresponding class, and a negative value indicates the opposite. The prediction from an example $\mathbf{x}$ to its complete label vector, $\mathbf{z}$, is the main task of multi-label learning [3].

The majority of previous multi-label learning methods assume that each example in the training set is associated with a complete label assignment. However, it is difficult to obtain a complete label assignment for each training example. For example, when the size of the candidate classes is large, such as in image annotation (see Fig. 1), it is costly to acquire the complete label vector for even one training image. Another possible reason of this difficulty is the ambiguity among classes, such as in AU recognition. AU is typically labeled by trained experts, which could be a time-consuming process. Furthermore, due to the ambiguity among AUs such as cheek raiser (AU6) vs. eye lid tightener (AU7) (see Fig. 2) as well as the poor image quality, some AUs are difficult to label confidently (the detailed definitions of all AUs can be found in [2]). A more realistic scenario is that each image is only provided with a partial label assignment, while the assignments with other classes are missing (see Fig. 1). To explicitly accommodate the missing labels, for each input image we introduce the definition of an incomplete label vector $\mathbf{y} \in \{-1, 0, 1\}^m$: a nonzero element has the same meaning as in the complete label vector $\mathbf{z}$, while a zero element corresponds to a label with no assigned value for this image, i.e., *missing label*.

The focus of this work is *multi-label learning with missing labels* (MLML) of which the primary objective is to derive a parametric multi-label classifier from input data example $\mathbf{x}$ and its incomplete label vector $\mathbf{y}$. Many previous multi-label learning methods that do handle missing labels usually do not make a clear distinction between missing labels and negative labels, i.e., the missing labels are assigned to the negative value by default [4–6]. However, such a simple treatment is often ungrounded in the practical applications. Consider the image annotation example in Fig. 1, label "home" is not shown in the label list of the image on the left, but we cannot simply conclude that the image should have a negative value for that label, because it has a positive value in the available label "house" that has strong semantic correlation with "home".

We formulate MLML as an inductive learning problem, which consists of three terms, including the consistency between the predicted labels and the provided labels, the smoothness of label

* Corresponding author. Tel.: +1 518 2766313.
E-mail addresses: wubaoyuan1987@gmail.com (B. Wu), lsw@cs.albany.edu (S. Lyu), bghu@nlpr.ia.ac.cn (B.-G. Hu), qji@ecse.rpi.edu (Q. Ji).

http://dx.doi.org/10.1016/j.patcog.2015.01.022
0031-3203/

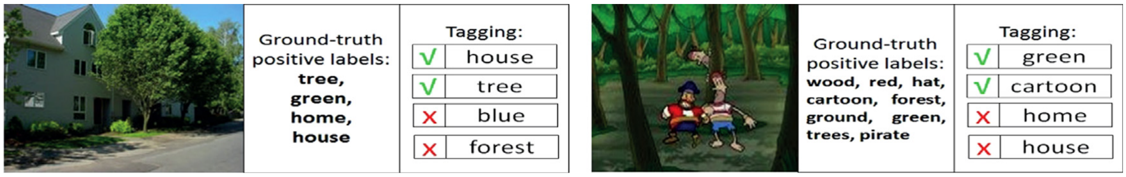

Fig. 1. Two examples of image annotation from the ESP Game data [8]. The symbol '✓' denotes the positive label, while '×' indicates the negative label. The labels not in the tagging list are missing labels. Take the left image as an example, "house" and "tree" are its positive labels, while "blue" and "forest" are negative labels. All other labels are missing labels. Note that in missing labels, some are actually positive labels, such as "green" and "home", as well as some negative labels, such as "cartoon" and "hat" (see the right image). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

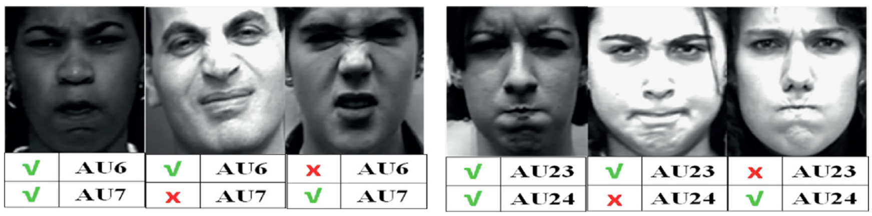

Fig. 2. An illustration of the ambiguity among AUs [2,9]. Under each facial image, its ground-truth AU labels are annotated (only the AU labels of interest are shown, while other AU labels are ignored): '✓' indicates the positive label, while '×' denotes the negative label. Two groups show different ambiguities: (left) the ambiguity between AU6 (cheek raiser) and AU7 (eye lid tightener); (right) the ambiguity between AU23 (lip tightener) and AU24 (lip presser). Due to such ambiguities, some facial images in CK+ database [9] are not labeled with respect to some AU classes.

assignments measured with similarities between examples and between classes, as well as a ℓ_{21} norm over the model parameter to avoid over-fitting. Our algorithm is obtained from the iteratively re-weighted least squares (IRWLS) [7] method. We demonstrate the superior performance of our method on two computer vision problems: one is the annotation of general images, and the other is the recognition of facial action unit (AU) on facial images. The main contributions of our work can be summarized as follows:

(i) we describe a new method to handle missing labels in multi-label learning;
(ii) we present an efficient numerical algorithm to learn the inductive classifier using the MLML framework;
(iii) we formulate image annotation and AU classification as multi-label learning with missing label problems, and show that our method outperforms several state-of-the-art works on three benchmark data sets.

2. Related works

In the literature of multi-label learning, some recent works have studied the problem with missing labels. The semi-supervised multi-label learning (SMSE2) [6] addresses the special case with examples either fully labeled or completely unlabeled. The weak label learning (WELL) method [4] focuses on the case where examples only have a partial set of positive labels available with the rest of the labels unassigned. Both SMSE2 and WELL assign the missing label values to zero by default, which coincides with the numerical value assigned to negative labels. As such, these methods implicitly assume that missing labels are equivalent to negative labels. This assumption is made explicitly in the work of multi-label learning with incomplete class assignment (MLR-GL) [5], where available labels all take positive values, and the missing labels are assigned to negative values, and thus becomes a fully labeled multi-label learning problem. However, as we have argued in Introduction, it is questionable if such an assumption always holds in actual data sets, and whenever it does not, treating missing label indiscriminately as negative labels introduce undesirable bias to the learning problem. The recent work of multi-label learning based on Bayesian compressed sensing (BML-CS) [10] can be used to solve the MLML problem, but it assumes a continuous probability model over the binary labels, and the resulting solution is based on a more costly MCMC algorithm. Moreover, the labels are assumed to be independently distributed in BML-CS, while the proposed model naturally incorporates the correlations between examples and between classes. In another related literature, i.e., matrix completion (MC), some recent works [11–13] have been proposed to handle the MLML problem. Their basic idea is concatenating the label matrix and feature matrix into a unified matrix, based on which the standard matrix completion techniques can be applied to fill in the missing labels. These works also avoid the label bias. However, the low rank assumption in MC implicitly implies class correlations. In contrast, our smoothness assumption explicitly captures such correlations, and we can also replace the smoothness to embed other types of correlations. Besides, the proposed efficient solution only involves matrix multiplication, while the solution to MC exploits the expensive SVD decomposition.

In the literature of image annotation, in addition to the single label methods [14–17], some multi-label learning methods have also been developed, such as [18–24]. However, a common assumption in these methods is that a complete label assignment for each training image should be provided. A few works also consider the case of incomplete labels of training images, such as WELL [4] and MLR-GL [5], which have been mentioned in the first paragraph of this section.

In AU recognition, some existing methods train a binary model with respect to each single class, such as [25–27,9]. The drawback of these methods is the ignorance of the relationships among different AU classes. Some recent models are developed to embed the spatial or temporal relationships among AUs, such as [28–33].

A similar learning problem with MLML, called multi-instance multi-label problem (MIML), has also been applied in image annotation [34–36] and scene classification [37,38]. Both MIML and MLML consider multi-label problems, but there exist two significant differences between them. For clarity, we take the image classification as the example. Firstly, their tasks are different. MIML aims to predict both image-level and region-level labels, while MLML only focuses on image-level label prediction. Secondly, in terms of predicting the image-level labels, MIML can be seen as supervised learning, while MLML can handle the missing labels.

3. Multi-label learning with missing labels

We start with a general setting of the MLML problem. We assume that the data are represented as matrix $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where each