Neurocomputing 157 (2015) 280–289

Contents lists available at ScienceDirect

Neurocomputing

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

# Multi-label learning with prior knowledge for facial expression analysis

Kaili Zhao [a,*], Honggang Zhang [a], Zhanyu Ma [a], Yi-Zhe Song [b], Jun Guo [a]

[a] School of Communication and Information Engineering, Beijing University of Posts and Telecommunications, Beijing, China
[b] School of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:
Received 26 April 2014
Received in revised form
1 January 2015
Accepted 7 January 2015
Communicated by Yongzhen Huang
Available online 21 January 2015

Keywords:
Facial expression analysis
Multi-label classification
Multi-label regression
Group Lasso
Total Variation

ABSTRACT

Facial expression is one of the most expressive ways to display human emotions. Facial expression analysis (FEA) has been broadly studied in the past decades. In our daily life, few of the facial expressions are exactly one of the predefined affective states but are blends of several basic expressions. Even though the concept of 'blended emotions' has been proposed years ago, most researchers did not deal with FEA as a multiple outputs problem yet. In this paper, multi-label learning algorithm for FEA is proposed to solve this problem. Firstly, to depict facial expressions more effectively, we model FEA as a multi-label problem, which depicts all facial expressions with multiple continuous values and labels of predefined affective states. Secondly, in order to model FEA jointly with multiple outputs, multi-label Group Lasso regularized maximum margin classifier (GLMM) and Group Lasso regularized regression (GLR) algorithms are proposed which can analyze all facial expressions at one time instead of modeling as a binary learning problem. Thirdly, to improve the effectiveness of our proposed model used in video sequences, GLR is further extended to be a Total Variation and Group Lasso based regression model (GLTV) which adds a prior term (Total Variation term) in the original model. JAFFE dataset and the extended Cohn Kanade (CK+) dataset have been used to verify the superior performance of our approaches with common used criterions in multi-label classification and regression realms.


## 1. Introduction

In recent years there has been a growing interest in improving all aspects of the interaction between humans and computers. As an essential way for human communications, facial expressions deliver rich information about human emotions. Facial expressions are most commonly used in everyday human-to-human communication, such as one smiles to show greeting, frowns when confused, and opens their mouths when surprised. The fact that we understand emotions and know how to react to other people's expressions greatly enriches the interaction. Researchers have tried to analyze facial expressions in an attempt to understand and categorize these expressions.

However, most previous attempts depict each facial picture with one of the predefined affective labels, such as the 6 basic affective states of happiness, sadness, surprise, fear, anger, and disgust [1]. They assume that each facial picture is linked to only one affective label, which tends to be over simplified. In common sense, people can express blended emotions. For example, when someone gets an unexpected birthday present from his best friend, he would be both happy and surprised. Few of the expressions are exactly one single predefined affective states (e.g. 100% happiness). Through surveys on blended emotions, we will illustrate reasons in Section 3 why we model FEA as a multi-label problem. Furthermore, most works in FEA field do not jointly model the continuous facial expression targets. That is to say, they treat each expression separately as a binary problem. Based on two points above, we model facial expression analysis as a multi-label problem at first, which depicts each facial expression with multiple continuous values and labels of predefined affective states. Then, multi-label learning algorithms are proposed jointly to model FEA. Algorithms are briefly introduced below.

First, a multi-label classifier is depicted. A number of algorithms have already been proposed for multi-label classification, such as multi-label KNN (ML-KNN) [2] and multi-label RBF (ML-RBF) [3]. However, in the experiment of multi-label FEA, they failed to show satisfied results since they did not describe the special relationship between different labels in FEA applications. Thus, it is necessary to analyze the relationship between different labels in FEA problem and design an ad hoc classifier. In past decades, there are many works related with AUs (active units) in FEA which describe facial muscle distributions when people show expressions. However, most previous research [4] depict that each facial picture should be linked with one of the predefined affective labels, such as the six basic affective states of happiness, sadness, surprise, fear, anger, and disgust. Based on this, we focus on a domain specific knowledge

* Corresponding author.
E-mail addresses: zhaokailibupt@gmail.com (K. Zhao), zhhg@bupt.edu.cn (H. Zhang), mazhanyu@bupt.edu.cn (Z. Ma), yizhe.song@eecs.qmul.ac.uk (Y.-Z. Song), guojun@bupt.edu.cn (J. Guo).

http://dx.doi.org/10.1016/j.neucom.2015.01.005
0925-2312/

that there are some common affective areas which relate to distributions of facial muscles in different expressions. So with this prior knowledge, we propose a multi-label learning algorithm with Group Lasso which groups the same affective areas (patches) of different expressions together. Then weights in one group will tend to be zero and non-zero simultaneously. Specially, some affective areas (patches) are depicted by super-pixels (resized images) and eigen-faces obtained by PCA. A Group Lasso regularized maximum margin (GLMM) algorithm is presented as a multi-label classifier to solve the proposed multi-label FEA problem above.

Second, based on the same physical meaning of Group Lasso above, multi-label Group Lasso regularized regression (GLR) is proposed to treat each facial expression output as continuous values. In order to verify the effectiveness of Group Lasso regularized term, support vector regression and $L_1$ norm regularized term based regression are used for comparison purpose. FEA has two schemes: static images based and videos based. Both of the proposed GLMM and GLR are good in static images and video sequences. In order to get better performance in video sequences, an improved regression model is constructed by taking consideration of the prior knowledge of continuous expressions changing of the same person in one video. Total Variation is a good regularized term to apply this domain specified knowledge. Thus GLTV (Group Lasso and Total Variation regularized regression) is proposed to analyze facial expressions specifically from video sequences.

At last, in order to make the experimental environment stable and consistent, a complete facial expression system is set up. We extract faces and align them from images or frames through RASL [5], which is robust for pictures with different illuminations, poses, and expressions. Furthermore, we choose PCA and resized gray-scale pixels as facial affective areas (patches). Then multi-label classification algorithm and regression models are implemented separately with corresponding labels. This paper introduces the following key contributions:

- Model FEA as a multi-label problem with six basic facial expression categories.
- Treat all facial expressions simultaneously when constructing FEA model.
- Propose multi-label classification and regression algorithms applied in both static images and video sequences. Depict two domain specified knowledge: the relationship between different expressions and the continuity of expressions between adjacent frames in one video.
- Extend and re-label CK+ dataset to a multiple outputs dataset with continuous values.

To start with, Section 2 illustrates related works of this paper. Section 3 describes the reason why FEA should be modeled as a multi-label problem. In Section 4.1, GLMM is proposed for multi-label facial expression classification and is compared with several state-of-art multi-label algorithms based on JAFFE dataset. Next, in order to illustrate the physical meaning of Group Lasso used in FEA more clearly, GLR is proposed as a regression problem in Section 4.2, and is also compared with several classical regression algorithms in JAFFE dataset. Section 4.3 further extends GLR into a more effective regression model (GLTV) applied in video sequences specifically. Section 5 draws the conclusion and points out future works of our proposed algorithms.

## 2. Related works

### 2.1. Facial expression analysis

A lot of attention has been paid to facial expression analysis (FEA), as it plays a very important role in human interaction. There are some interesting applications such as virtual reality, video conferencing, customer satisfaction studies for broadcast and web services, and some smart environments requiring efficient facial expression analysis. In industrial areas, there are also some existing FEA applications: Computer Expression Recognition Toolbox (CERT)[1] developed by Emotient's Founders, FEA system produced by Noldus[2] and so on.

According to researches in psychology, the main approach of emotion modeling is category-based method [6]. In this method, six basic emotions (happiness, sadness, surprise, fear, anger and disgust) are confirmed and these labels become the most universal targets in FEA. In this paper, we design models based on these basic emotions.

Learning based method is very popular in the realm of FEA with the above basic emotions as targets. There are some researchers focusing on regression model with different special targets in FEA. Rudovic et al. [7,8] proposed a Gaussian Process Regression model to learn the mappings between non-frontal and frontal poses. Bilinear Kernel Reduced Rank Regression was proposed in paper [9] to learn high-dimensional mapping between face appearance and expressions. In the meantime, features learning methods also have been studied. For example, Zhang et al. [10] extracted useful features by Monte Carlo algorithm to better represent facial expressions.

However, there are two drawbacks should be mentioned . First, one common concern of previous approaches is the assumption that each facial picture is linked to only one affective label, which tends to be over simplified. Second, current works in FEA field mostly do not jointly model facial expressions targets, that is to say, most of them treat each expression separately as a binary problem. The proposed algorithms in this paper overcome these two problems by modeling multi-label learning methods based on Group Lasso [11,12].

### 2.2. Group Lasso

$l_1$ regularization [13] is a frequently adopted technique in machine learning. However, $l_1$ norm itself cannot incorporate the structure information between multiple labels. Group Lasso overcomes this problem [12] and can be expressed as follows:

$$\hat{\boldsymbol{\beta}}^{GL} = \arg\min_{\beta}\left(\mathrm{Loss}(\beta)+\lambda\sum_{j=1}^{J}\|\beta_j\|_{l_2}\right). \tag{1}$$

With Group Lasso regularization, the parameters $\boldsymbol{\beta}$ are assumed to be clustered into groups. Instead of summing the absolute values of each parameter, the sum of Euclidean norms of the parameters in each group is used. Intuitively, this should drive all the weights in one group to zero or non-zero simultaneously, thus leading to group selection [12].

The penalty function used in Eq. (1) is intermediate between the $l_1$ penalty used in the Lasso [13] and $l_2$ penalty used in Support Vector regression [14]. Thus Group Lasso term encourages sparsity in individual coefficients between groups and treats directions in the same group equally while $l_2$ penalty treats all directions equally. As Fig. 1 depicted, consider a case in which there are two factors and the corresponding coefficients are a 2D vector $\boldsymbol{\beta}_1=(\beta_{11},\beta_{12})'$ and a scalar $\boldsymbol{\beta}_2$. $\beta_{11}$ and $\beta_{12}$ are placed in one group and $\boldsymbol{\beta}_2$ is the other one. So $\beta_{11}$ and $\beta_{12}$ will vary equally in $l_2$ and be sparsity as $l_1$ at the same time.

[1] http://www.emotient.com/cert

[2] http://www.noldus.com/facereader/facial-expression-analysis