ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Cumulative attribute relation regularization learning for human age estimation

Qing Tian, Songcan Chen *

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

A R T I C L E I N F O

Article history:
Received 14 November 2014
Received in revised form
13 January 2015
Accepted 15 March 2015
Communicated by Dr. Qingshan Liu

Keywords:
Age estimation
Cumulative attribute
Mutual relation
Regularization
LS-SVR
Ridge regression

A B S T R A C T

In recent years, the problem of human face-based age estimation has attracted increasing attention due to its extensive applicability and motivated a variety of approaches being proposed, in which the method based on the coding of *cumulative attribute* (CA) achieves competitive performance by taking into account both the neighbor-similar and the ordinal characteristics of ages. However, in their learning, the inherent *mutual relations* between the CA codes have not been exploited, thus leaving us a performance space that can be improved. To this end, in this work we first derive such relations by performing the difference-like operation between the CA codes in certain order to construct so-called 0-order and 1-order relation matrices and then incorporate them as two corresponding regularization terms, coined as CA-oriented ordinal structure regularization (CAOSR) and CA-oriented adjacent difference orthogonal regularization (CAADOR), into the objective of the multi-output regressor. Consequently, corresponding CA-based regressors regularized with the mutual relations are developed. Finally, through extensive experiments on three human aging datasets, the FG-NET and the Morph Album 1 and Album 2, we demonstrate the effectiveness of our strategies in improving CA-based age estimation.

1. Introduction

In machine learning, a large number of problems are related to human face due to that rich information is contained in it, such as facial expression, gender, race and age, in which the problem of human face-based age estimation has aroused increasing attention due to its wide applications such as web security control [12,17], ancillary identity authentication [14], and advertisement recommendation [27], etc.

In order to conduct age estimation based on human face, a variety of approaches have been proposed to date. Generally, they fall into three categories: *classification-based*, e.g., [17,9,30,1,28], *regression-based*, e.g., [18,7,22,33,32,10,4,19,20], and *their hybrid*, e.g., [12,13,16].

When we consider each age as a separate class, the age estimation can be made under ordinary classification framework. For example, Lanitis et al. [17] extracted AAM features from facial images and respectively applied the nearest neighbor classifier and artificial neural networks for age estimation and achieved comparable performance. Geng et al. [9] specially designed a three-layer conditional probability neural network (CPNN) to capture the age contribution information for age classification. Moreover, Ueki et al. [30] conducted age group classification by building Gaussian mixture models after discriminative dimensionality-reduction and received promising results respectively for male and female on several famous age datasets. More recently, Alnajar et al. [1] employed the soft coding to extract codebooks for age group classification and received better estimation on an unconstrained real-life dataset than the hard coding approaches. And Sai et al. [28] even used the extreme learning machines to perform age group estimation and obtained competitive results.

Actually, the age estimation is more of a regression problem than a generic multi-class classification due to the continuity of aging. According to this characteristic, many attempts have been made. For instance, Lanitis et al. [18] established a quadratic function to fit the ages with facial images represented by AAM features. Fu et al. [7] borrowed the multiple linear regression to learn an aging prediction function in the manifold space. And Luu et al. [22] employed the off-the-shelf ξ-SVR [31] for aging function learning. Moreover, to handle the uncertainty of annotations of age labels, Yan et al. [33] constructed a semi-definite programming (SDP) regression model to train an aging regressor. Although the SDP regressor can relatively model the age labels' uncertainty better, the learning is very time-consuming. To reduce the time complexity, they [32] then proposed to speed up the SDP learning by using the Expectation–Maximization (EM). Furthermore, Geng et al. [10] proposed the aging pattern regressing (AGES) to

* Corresponding author.
E-mail addresses: tianqing@nuaa.edu.cn (Q. Tian), s.chen@nuaa.edu.cn (S. Chen).

http://dx.doi.org/10.1016/j.neucom.2015.03.078
0925-2312/

Please cite this article as: Q. Tian, S. Chen, Cumulative attribute relation regularization learning for human age estimation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.078

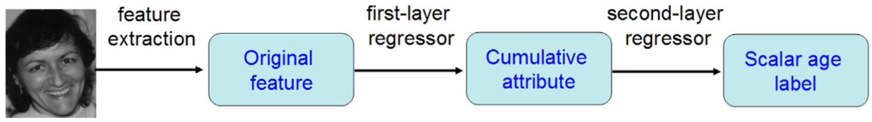

Fig. 1. The flowchart of CA-based human age estimation.

generate age labels for missing patterns. Although the methods aforementioned can yield age estimation performance to different extents, they ignored the fact that there exists natural ordinality among ages [5,4]. To this end, Chang et al. [4] specially designed an ordinal hyperplanes ranker (OHRank) for age estimation and on FG-NET dataset they obtained better performance than AGES. Later, Li et al. [19] presented a distance-based ordinal regressor for age estimation, in which the ordinal information of ages is incorporated into the metric and on FG-NET they obtained competitive performance. Moreover, they [20] took the ordinality and local manifold structure preserving ability as a criterion to perform feature selection and conducted age regression with much competitive results. More recently, they [21] presented an ordinal metric learning method for image ranking by preserving both the local geometry information and the ordinal relationship of the data.

Although the methods reviewed above can perform encouraging human age estimation with different performance, they have not exploited another essential characteristic of the ages that neighboring ages are generally more similar in facial appearance than those apart. For example, the facial appearance of 11-year-old is more similar to that of 13 compared to that of 30, as exhibited in Fig. 2 (in Section 2). This characteristic is of help in estimating the ages, especially when the age distribution is imbalanced [5], because similar ages can be used to partially depict their neighboring ages that are absent in the learning and thus alleviate the imbalance. Therefore, such *neighbor-similarity* of ages should also be incorporated into the estimation. To simultaneously consider both the *ordinality* and the *neighbor-similarity* of the ages,[1] Chen et al. [5] proposed the *cumulative attribute* (CA) coding to represent the age. Concretely, they first used the multivariate ridge regression (mRR) [2] to transform the instance from its original input feature to a CA code; and then applied a second-layer scalar-output regressor to map the CA code to a scalar age label. The flowchart of the two-layer regression is shown in Fig. 1, and by this way they obtained competitive age estimations.

Although the characteristics of ordinality and neighbor-similarity of the ages are considered in the CA coding, the inherent *mutual relations* explicitly or implicitly existing between the CA codes have not been exploited for learning, thus leaving us a room of promoting its performance. To this end, in this work we first derive such relations by performing difference-like operations on the *CA coding matrix*[2] to construct so-called 0-order and 1-order *relation matrices*,[3] respectively. Then, we formulate the *relation matrices* as two corresponding regularization terms, coined as *CA-oriented ordinal structure regularization* (CAOSR) and *CA-oriented adjacent difference orthogonal regularization* (CAADOR). And, in order to take the *mutual relations* into the CA learning, we regularize the first-layer regressor (as shown in Fig. 1) by embedding the regularization terms, CAOSR and CAADOR, into its objective. Finally, through extensive experiments, we demonstrate the effectiveness of our strategies in improving CA learning on human age estimation.

The rest of this paper is organized as follows. In Section 2, we briefly review related work on CA coding. In Section 3, we derive two types of regularization terms, coined as *CAOSR* and *CAADOR*, to depict the mutual relations among the CA codes, and embed them into the objectives of the mRR and mLS-SVR, both of which act as the first-layer regressor in the CA learning, in Section 4. In Section 5, we conduct experiments to evaluate our strategies. Finally, we conclude the paper in Section 6.

2. Related work

Following the spirit of literature such as [23], Chen et al. [5] presented the *cumulative attribute*(CA) coding for learning in such scenarios as human age estimation. Concretely, given a set of N training samples $\{x_i, l_i\} \in \Re^D \times \Re$, $l_i \in \{1, 2, \ldots, K\}$, $i=1,2,\ldots,N$, where x_i denotes the ith instance and l_i is its corresponding scalar label, D denotes the feature dimensionality of x_i and K is the number of classes (e.g., the scale of the aging range). Here for the ith sample x_i, its scalar label value l_i, e.g., the age value, is transformed into a K-dimensional vector y_i, named as *cumulative attribute* (CA) code, whose jth element is defined as

$$y_i^j = \begin{cases} 1, & j \le l_i \\ 0, & j > l_i \end{cases}$$

where $j=1,2,\ldots,K$. As a comparison, the *non-cumulative attribute* (NCA) is given as well with the jth element defined as

$$y_i^j = \begin{cases} 1, & j = l_i \\ 0, & j \ne l_i \end{cases}$$

As argued for age estimation, the CA coding can relatively well capture the characteristic that the attribute values at neighboring ages should be more similar than those further apart. Moreover, it can alleviate the challenge of the insufficient and imbalanced sample distribution within the entire aging range, while the NCA coding cannot. The appealing characteristic of CA coding can be intuitively demonstrated in Fig. 2.

Just as shown in Fig. 2, the facial age appearances between (a) and (b) obviously are more similar than those between (a) and (c), which is consistent with the coding differences between their corresponding CA, i.e., the difference of 2 years between (a) and (b) is smaller than that of 19 between (a) and (c). However, using NCA, the age difference between (a) and (b) and that between (a) and (c) are the same and both equal to 1, by which the similarity of neighboring ages is not reflected at all, according to its definition above. Therefore, the coding way of CA is reasonable and desirable to depict human facial age.

Now given a human facial image, x_i, in order to learn a mapping $y_i = W^T x_i + B$ from its original feature x_i to the CA representation y_i, where $W = [w_1, \ldots, w_K] \in \Re^{D \times K}$ is the weight matrix and $B \in \Re^K$ is the bias, Chen et al. [5] used the mRR because of its robustness in regression, formulated as

$$\min_{W,B} \frac{1}{2} \sum_{i=1}^{N} \|y_i - (W^T x_i + B)\|_F^2 + \frac{\lambda}{2} \|W\|_F^2, \quad (1)$$

[1] Note the difference between the ordinality and the neighbor-similarity of ages. The ordinality defines the global order relationship of the ages while the neighbor-similarity states the similarity of facial appearances at close ages. Moreover, the neighbor-similarity is also conceptually different from the local manifold geometry relationships [20,21] of the ages. The local manifold geometry structures describe the neighbor relationship of ages in manifold space, while the neighbor-similarity stated in this paper depicts the biological similarity of facial appearances of neighboring ages.

[2] The *CA coding matrix* refers to such a matrix in which each column corresponds to a CA code for an instance from some class, as shown in Fig. 3(a).

[3] For why just extracting the 0-order and 1-order relation matrices, please refer to the *Remark* in Section 3.

Please cite this article as: Q. Tian, S. Chen, Cumulative attribute relation regularization learning for human age estimation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.03.078