Learning Distance Metric Regression for Facial Age Estimation

Changsheng Li[1], Qingshan Liu[1,2], Jing Liu[1], Hanqing Lu[1]
[1]*NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190*
[2]*CICE, Nanjing University of Information Science and Technology, Nanjing, China, 210044*
{*csli,qsliu,jliu,luhq*}*@nlpr.ia.ac.cn*

Abstract

This paper proposes a novel regression method based on distance metric learning for human age estimation. We take age estimation as a problem of distance-based ordinal regression, in which age difference is measured by an efficient distance metric. To reach this goal, we propose to learn such a distance metric that can preserve both the ordinal information of different age groups and the local geometry structure of the target neighborhoods simultaneously. Then, the facial aging trend can be truly discovered by the learned metric. Experimental results on the publicly available FG-NET database are very competitive against the state-of-the-art methods.

1. Introduction

In recent years, human age estimation attracted much attention in computer vision and pattern recognition [3, 2, 6, 7]. The purpose of age estimation is to label a face image automatically with the exact age (year) or the age group (year range). To achieve this goal, many algorithms formulate age estimation as a regression problem [6, 14, 4], in which the age labels are taken as numerical values.

However, it should be noted that the performance of these regression models is dependent critically on their being given distance metric over the input space. Since aging face process is an ordinal procedure in temporal domain, the faces with smaller age difference should be more related than the faces with larger age difference. In a sense, it is similar to learn an appropriate distance metric to measure the distance between the aging faces, so that the distance between the face representations with smaller age difference should be shorter than the distance between face representations with larger age difference.

Generally, distance metric learning is to learn a distance metric for an input space from some side information such as must-link/cannot-link constraints between data instances. The optimal distance metric is found such that the objects in must-link constraints are close to each other while the objects in the cannot-link constraints are well separated. So far, various distance metric learning algorithms have been proposed, such as the convex programming approach [13], relevance component analysis [8], neighborhood component analysis (N-CA) [5], metric learning via large margin nearest neighbor [11], and Bayesian distance metric learning [15].

Almost all the proposed distance metric learning techniques are only applicable for the tasks of classification and clustering, they are not focusing on regression problem. Taking age estimation as an example, must-link/cannot-link constraints in the distance metric learning algorithms can make the distances among the faces representation with the same age label minimized and the distances among the faces representation with different age labels maximized, but it cannot preserve the ordinal information among aging faces, which is of great importance in age estimation. Recently, [12] proposed a distance metric learning approach for regression and applied to age estimation. It aimed to preserve the local geometry of the same semantic neighborhoods, but it did not take account of keeping the ordinal relationship.

In this paper, we propose a novel algorithm to learn distance metric for regression (LDMR), and apply it for age estimation. The proposed approach targets at: 1) maximizing the margins between two aging face groups with different age labels; 2) keeping the ordinal information among aging faces groups of different age labels; 3) preserving the local geometry of the target neighborhoods. Mathematically, we formulate it as a semidefinite programming (SDP) problem. Based on the learned metric, we build a new kNN regression model for age estimation. Experimental results on FG-NET Aging Database [1] demonstrate its power compared to the state-of-the-arts.

2. Formatting your paper

Let $\{(\mathbf{x}_i, y_i)\}(i = 1, 2, \cdots, n)$ denote a training set of n labeled examples with inputs $\mathbf{x}_i \in \mathbb{R}^d$ and their associated continuous non-negative age labels y_i. The goal is to learn a Mahalanobis metric $\mathbf{A}$ that defines distances of the form:

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_{ij}\|_{\mathbf{A}} = \sqrt{\mathbf{x}_{ij}^T \mathbf{A} \mathbf{x}_{ij}} \quad (1)$$

where $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, and $\mathbf{A}$ is a parameter in the distance function that is a positive semi-definite matrix.

In [12], the authors try to preserve the local geometry of the same semantic neighborhoods. An example is given by Figure 1(a). Before training, the five points nearly have the identical distances to $\mathbf{x}_i$ in the original 2-D space as shown in the left part of Figure 1(a). After training, the three yellow points have the same relative distances to $\mathbf{x}_i$. Meanwhile, the red point and the blue point are separated from $\mathbf{x}_i$ as far as possible, which is shown in the right part of Figure 1(a). However, the ordinal information among all points is not well preserved, which may reduce the regression performance.

The desired result should be like Figure 1(b), in which the ordinal information is well kept according to the learned distance metric. To reach this goal, we formulate the objective function as follows:

$$\max\ J(\mathbf{A}) = \sum_{i,j} w(i,j) d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) \quad (2)$$

$$s.t.\ \ d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = d^2_{ij} \quad \text{if } \eta_{ij} = 1 \quad (3)$$

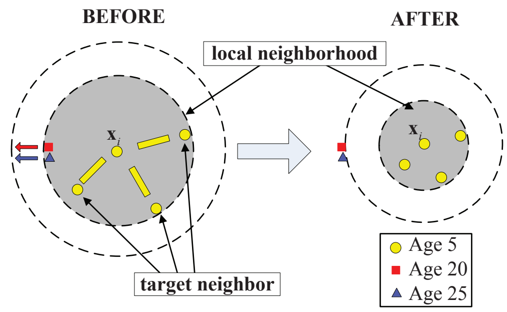
$$\mathbf{A} \succeq 0 \quad (4)$$

where $w(i,j)$ is a weighting factor, and is defined as:

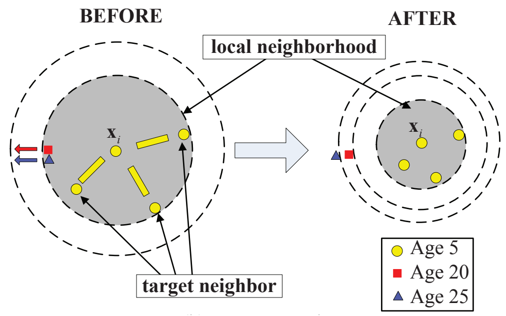
$$w(i,j) = \begin{cases} (\frac{D(i,j)+\delta}{C-D(i,j)})^m & \text{if } y_i \neq y_j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $D(i,j)$ is the absolute label difference between $\mathbf{x}_i$ and $\mathbf{x}_j$. δ refers to the labeling noise. $C = D(i,j) + \theta, \theta > 0$ which ensures the denominator not to be zero. m is chosen to make data easier to discriminate. In addition, $\eta_{ij} \in \{0, 1\}$ indicates whether $\mathbf{x}_j$ is $\mathbf{x}_i$'s one of K target neighbors, i.e., K other inputs with the same age label y_i having the minimal distance to $\mathbf{x}_i$. d_{ij} is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$.

In Eq.(2), the objective function penalizes small distances between each input and all other inputs without sharing the same age label. More importantly, the weighting factor $w(i,j)$ makes the objective function penalize small distances between two inputs with large age difference, which can preserve the ordinal information of different age groups. The constraint condition



(a) The strategy of [12]



(b) Our approach

Figure 1. Schematic illustration of the metric learning approach. Here we only take the sample $\mathbf{x}_i$'s neighborhoods as an example. In both Figure (a) and (b), the left part is the situation of neighborhood before training and the right part delineates the situation after training. Data with the same age labels are marked in the same shape and color.

makes sure that the distance between the target neighborhoods is unchanged, which can preserve the local geometry of the target neighborhoods. In the absence of prior knowledge, we use Euclidean distance to represent the distance between the target neighborhoods in the original space.

In practical applications, not all constraints in Eq.(3) could be satisfied, therefore we add some slack variables to ensure that the feasible is not empty as follows:

$$\max\ J(\mathbf{A}) = \sum_{i,j} w(i,j) d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) + \alpha \sum_{l} \epsilon_l \quad (6)$$

$$s.t.\ \ d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) + \epsilon_l = d^2_{ij} \quad \text{if } \eta_{ij} = 1 \quad (7)$$

$$\mathbf{A} \succeq 0 \quad (8)$$

$$\epsilon_l \geq 0 \quad (9)$$

This problem is a standard semidefinite programming problem, and it can be effectively solved to obtain its global maximum by several online generic packages [9].

After obtaining the learned distance metric $\mathbf{A}$, we build a new kNN regression model as follows:

$$\widehat{y} = \frac{\sum_{i=1}^{K} \lambda_i \Lambda_i}{\sum_{i=1}^{K} \lambda_i} \tag{10}$$

where $\widehat{y}$ is the estimated age label of the testing sample $\widehat{\mathbf{x}}$,$\{\Lambda_i, i = 1, 2, \cdots, K\}$ are the corresponding age labels of $\widehat{\mathbf{x}}$'s K nearest neighbors in the training set and λ_i is the weighting coefficient, which is calculated as:

$$\lambda_i = \frac{1}{d_{\mathbf{A}}^2(\widehat{\mathbf{x}}, \mathbf{x}_i)} \tag{11}$$

3. Experiment

We evaluate the performance of the proposed method (LDMR) on the public available FG-NET database [1]. This database contains totally 1002 color or gray images from 82 different persons with the age ranges from 0 to 69 years. We use AAM [3] as the feature extraction method in the experiments because it is capable of extracting both the shape and the appearance features of human images.

For all the experiments, we set $C = 80, \delta = 1, m = 0.4$, and the nearest neighborhood number $K = 20$ empirically. For all the statistical experiments, we follow the popular test scheme, namely Leave-One-Person-Out (LOPO), which is usually taken for the FG-NET database, as in [2, 6, 12, 4]. The mean absolute error (MAE, i.e., $\frac{1}{n}\sum_{i=1}^{n} |\widehat{y}_i - y_i|$) is used as the evaluation measure. Here $\{\widehat{y}_1, \ldots, \widehat{y}_n\}$ denote the estimated ages and $\{y_1, \ldots, y_n\}$ are the true ages.

3.1 Experimental Results

To evaluate the effectiveness of the learned metric, we compare it with two related distance metric learning approaches [12] and [11]. We also use the Euclidean metric as the baseline. For fair comparison, all the distance metrics are combined with kNN regression to do age estimation. Following [12], we first randomly choose 400 training samples from the whole training set to learn the distance metric, and then apply kNN regression to the whole training set for age estimation. Table 1 reports the results. It is clear to see LDMR outperforms the other distance metric learning methods.

Table 1. MAEs over different age ranges on FG-NET Database wrt kNN in different metrics.

Age Range	Image Num.	mkNN [12]	lmNN [11]	Eucli-Metric	LDMR
00-09	371	2.29	3.16	5.65	**2.06**
10-19	339	3.65	3.50	3.39	**3.09**
20-29	144	5.44	5.14	4.91	**3.61**
30-39	79	10.55	11.52	13.13	**9.16**
40-49	46	**15.81**	20.17	22.39	16.30
50-59	15	**25.18**	27.33	30.27	26.33
60-69	8	36.80	39.86	43.13	**34.5**
Overall	1002	4.93	5.66	6.83	**4.51**

We also compare LDMR with the existing results on the FG-NEG database reported in recent year. Table 2 lists the MAEs reported in the literatures together with that of our algorithm. LDMR outperforms nearly all the other methods except OHRank [2]. Although the MAE of OHRank is slightly smaller than that of LDMR, the implementation of OHRank is much harder and the computation cost is much higher. In LDMR, what we need is to learn a new distance metric, and then build a kNN regression on the learned metric. In OHRank, it needs to train multiple SVM classifiers for age estimation. LDMR is also competitive to [4], while [4] is based on a nonlinear distance metric learning method. Moreover, the amount of the training samples used in [4] for learning the metric is more than twice that of our work.

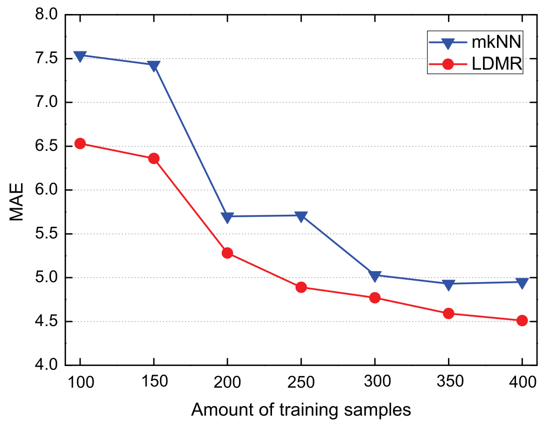


Figure 2. MAE of different methods using different amounts of training samples.

To further disclose the relationship between the performance and the amount of training data used for learning distance metric, we test the MAE results of LDMR under different amounts of training samples ranging from 100 to 400. We compare it with mkNN [12], which has already shown its effectiveness in Table 1. The results are shown in Figure 2. The proposed distance metric outperforms mkNN under all differen-

Table 2. MAE comparisons of different algorithms

Method	MAE
LARR [6]	5.07
RPK [14]	4.95
BIF [7]	4.77
Nonliear metric+kNN [4]	4.67
OHRank [2]	4.48
LDMR	4.51

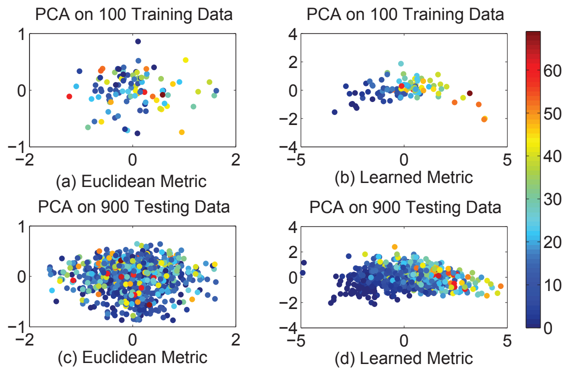


Figure 3. The 2-D visualization of PCA on the Euclidean metric and the learned metric. The color of the point represents age with blue being the youngest and red being the oldest.

t amounts of training samples.

To show the learned distance metric is helpful for discovering an apparent aging trend intuitively, we randomly choose 100 facial images as the training data to learn the distance metric and the rest of images are used as the testing data. Figure 3(a) and (b) give the details of the training stage. Figure 3(a) displays the 2-D visualizations of principal component analysis (PCA) [10] on the training data using Euclidean metric. Figure 3(b) shows the 2-D visualizations of PCA on the same training data using the learned metric. We can see the aging trend is clear by using our metric. Figure 3(c) and (d) displays the 2-D visualizations for the testing data. The aging trend is still discovered apparently by our learned metric.

4. Conclusion

In this paper we present a novel distance metric learning algorithm for age estimation. By keeping ordinal information of different age groups and preserving the local geometry of target neighborhoods, the learned metric can truly discover the aging trend. Experimental results on a widely used public aging database validate the effectiveness of the proposed method.

5. Acknowledgments

This work was supported by the National Natural Science Foundation of China 973 Program (Project No. 2010CB327905) and the National Natural Science Foundation of China (Grant No. 60903146, 60835002)

References

[1] The FG-NET aging Database, availabel at http://sting.cycollege.ac.cy/alanitis/fgnetaging/index.htm.

[2] K. Chang, C. Chen, and Y. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Proc. CVPR*, 2011.

[3] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *Proc. ECCV*, pages 484–498, 1998.

[4] N. Fan. Learning nonlinear distance functions using neural network for regression with application to robust human age estimation. In *Proc. ICCV*, 2011.

[5] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In *Proc. NIPS*, 2005.

[6] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans on Image Processing*, pages 1178–1188, 2008.

[7] G. Guo, G. Mu, Y. Fu, and T. Huang. Human age estimation using bio-inspired features. In *Proc. CVPR*, 2009.

[8] N. Shental, T. Hertz, D. Weinshall, and M. Pavel. Adjustment learning and relevant component analysis. In *Proc. ECCV*, 2002.

[9] J. F. Sturm. Using sedumi 1.02. a matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, pages 625–653, 1999.

[10] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. CVPR*, pages 586–591, 1991.

[11] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *Proc. NIPS*, 2006.

[12] B. Xiao, X. Yang, Y. Xu, and H. Zha. Learning distance metric for regression by semidefinite programming with application to human age estimation. In *Proc. ACM Multimedia*, 2009.

[13] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *Proc. NIPS*, 2002.

[14] S. Yan, X. Zhou, M. Liu, M. H. Johnson, and T. Huang. Regression from path-kernel. In *Proc. CVPR*, 2008.

[15] L. Yang, R. Jin, and R. Sukthankar. Bayesian active distance metric learning. 2007.