ISSN: 2321-7782 (Online)
Volume 2, Issue 9, September 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

An Efficient Method for Human Age Estimation by Label Distribution Learning

Vinod G. Khetade[1]
Computer Department
Rajarshi Shahu College of Engineering
Pune – India

S. B. Thakare[2]
Computer Department
Rajarshi Shahu College of Engineering
Pune – India

Abstract: Age information is useful in a variety of applications, such as human-computer interaction, surveillance monitoring, security control and video content analysis. Hence developing age estimation methods has become an interesting topic emerging in recent years. An efficient methods for facial age estimation based on label distribution learning is proposed in this paper. This method considers each face image as an instance associated with a label distribution. The label distribution covers a certain number of neighboring ages, representing the degree that each face age describes the facial appearance. As a result, label distribution will make a face image contribute to not only the learning of chronological age, but also the learning of its neighboring ages. The learning algorithms named IIS-LLD and modified CPNN for label distribution learning are proposed and applied to the problem of facial age estimation. Our experimental results show that proposed label distribution learning algorithm is better than previous methods.

Keywords: machine learning, label distribution, neural network, age estimation.

I. INTRODUCTION

A human face image contains huge information about personal characteristics, including identity, emotional expression, gender, pose, beards etc. Human face is robust because it changes in a short period. With the progress of aging, human faces shows remarkable changes such as face size getting larger, face skin becomes darker and wrinkly. The main goal of age estimation is to compute a person's exact age or age-group based on face attributes derived from a facial image. One of the main challenges of facial age estimation is the lack of sufficient training data. A suitable training data should include multiple face images of the same person covering a wide range of ages. Since the aging progress is uncontrollable, the collection of such a training dataset usually requires great efforts in searching past images and we cannot obtain future images. Consequently, the available training dataset just contain few face images of each person and images at higher ages are especially rare.

One of the main difficulties in age estimation is that the learning algorithms cannot expect sufficient training data. Fortunately, the faces at the neighboring ages look quite similar which results from the fact that aging is slow and gradual process. Multiple age numbers can be used to indicate appearance of one face. Inspired by this observation, the basic idea behind this paper is to use neighboring face images to learn a particular age. The utilization of adjacent ages is achieved by introducing a new labelling paradigm, i.e. assigning a label distribution to each image rather than a single label of the chronological age. A suitable label distribution will make a face image contribute to not only the learning of its chronological age, but also the learning of its adjacent ages [7]. Accordingly, a learning algorithm different from traditional learning schemes are proposed for learning from the label distributions.

Fig. 1 The aging faces of one subject in the FG-NET Database.
The chronological ages are given at the bottom

II. LITERATURE REVIEW

An early work on exact age estimation was reported by Lanitis et al., where the aging pattern was represented by a quadratic function called aging function, and the WAS (Weighted Appearance Specific) method [2] and AAS (Appearance and age specific) method [1] were proposed. Geng et al. [6] proposed the AGES algorithm based on the subspace trained on a data structure called aging pattern vector. The basic idea is to model the aging pattern, which is defined as the sequence of particular individuals face images sorted in time order, by constructing a representative subspace. Later, various methods have been developed for facial age estimation. For example, Fu et al. [8] [9] proposed an age estimation method based on multiple linear regressions. Guo et al. [3] designed a locally adjusted robust repressor to predict the human age by using support vector regression (SVR) method. Yan et al. [5] regarded age estimation as a regression problem with nonnegative label interval and solved the problem through semi definite programming (SDP). They also proposed EM (Expectation Maximization) algorithm to solve the regression problem and speed up the optimization process. Zhang and Yeung [10] proposed the multi-task warped Gaussian process to learn a separate age estimator for each person. Chang et al. [4] proposed the ordinal hyperplane ranking algorithm by using cost sensitive binary classification.

III. IMPLEMENTATION DETAIL

A. Label Distribution

Age is usually measured in years which are continuous time spectrum. Thus, the label distribution can be treated as a continuous distribution in this paper. In the label distribution, a real number $d_z^y \in [0, 1]$ is assigned to each label y called as description degree. This degree gives full class description of the instance. The addition of description degree of all the labels should be up to 1.There are three cases of label distribution, single, multiple labels, and general case of label distribution. Examples of typical label distributions for five class labels, for case 1, a single label is assigned to the instance, so $d_z^{y_2}$=1 means that label y_2 fully describes the instance. For case 2,two class labels y_2 and y_4 are assigned to the instance, so each label only describes 50 percent of the instance, i.e. $d_z^{y_2}$=$d_z^{y_4}$=0.5.Finally case 3 gives a general case of label distribution [7] which satisfies the constraint $d_z^y \in [0,1]$ and $\sum_y d_z^y$=1.

The Description degree d_z^y is not the probability that class y correctly labels the instance, but the degree that y describes the instance. Thus in the distribution, the correct labels are the labels with nonzero description degree. Label Distribution reflects the ambiguity of the class description of the instance, i.e., one class label may only partially describe the instance. The chronological age 35 and the neighboring ages 34 and 36 can be used to describe the face appearance of a 35 year old face. For each of 34, 35, and 36, it is completely true that it can be used to describe the face appearance. The description degree of each age's indicates how much the age contributes to the full class description of the face.

B. Learning from Label Distributions

In label distribution, we represent the description degree in the form of conditional probability, i.e., d_z^y = P(y|z). Then label distribution learning [7] can be defined by:

ISSN: 2321-7782 (Online)

Let Y= { $y_1, y_2,$ ……..,y_c } denotes the finite set of class labels. And Z is the set of instance. Given a training set S={$(z_1, D_1), (z_2, D_2)$……., (Z_n, D_n)}, where $z_i \in$ Z is an instance, D_i={ $d_z^{y_1}, d_z^{y_2}$,….. $d_{z_i}^{y_c}$ } is the distribution of the random variable y ∈ Y associated with z_i, and the goal is to learn a conditional probability from S, where z ∈ Z and y ∈ Y.

Suppose P(y|z) is a parametric model P(y|z; θ),where θ is the vector of the model parameters. Given a training set S, the goal of LLD is to find the θ, that can generate a distribution similar to D_i, given an instance z_i. The similarity between two distributions is measured by the Kullback-Leibler divergence.

C. Mathematical Model

INPUT: $S = \{(z_i, D_i)\}_{i=1}^{n}$ is the training set.

Where,

Y= { $y_1, y_2,$ ……..,y_c } is the set of aging face images of one subject.

Z= { $z_1, z_2 \ldots \ldots z_n$ } set of subjects.

D_i={ $d_z^{y_1}, d_z^{y_2}$,….. $d_{z_i}^{y_c}$ } set of description degrees.

$$d_z^y = \frac{P(y)}{\sum_y p(y)} \quad (1)$$

T(θ) is the target function.

$$T(\theta) = \sum_{i,j} d_{z_i}^{y_j} \log p(y_j|z_i;\ \theta) \quad (2)$$

OUTPUT : p(y|z ; θ)

D. The IIS-LLD Algorithm

There is a presumption in IIS-LLD [7], i.e., the conditional probability of an age given the face image is modeled by the maximum entropy model. However, the maximum entropy model may not fit the age estimation problem well. It is better to learn the model based on the training data. The Label Distribution Learning Algorithm [7] is described as follows:

1 Initialize the model parameter vector $\theta^{(0)}$;

2 i ← 0;

3 *repeat*

4 i ← i + 1;

5 Solve $\frac{\partial A(\Delta|\theta)}{\partial \delta_{yj,k}} = \sum_i d_{z_i}^{y_j} gk(z_i) -$

$$\sum_i p(y_j\ |z_i;\theta) g_k(z_i) \exp\left(\delta_{y_j,k}\ s(g_k(z_i)) g^{\#}(z_i)\right) = 0.$$

6 $\theta^{(i)} \leftarrow \theta^{(i-1)} +\ \Delta$;

7 *Until* $T(\theta^{(i)})$ - $T(\theta^{(i-1)}) < \epsilon$;

8 p(y|z ; θ) ← $\frac{1}{u}$ exp($\sum_k \theta^{(i)}{}_{y,k}\, gk(z)$);

E. Modified CPNN Learning Algorithm

The IIS-LLD Algorithm assumes the derivation of P(y|z) as the maximum entropy model. This algorithm removes this assumption by using a three layer neural network. We are updating weight of the neural network using forward-propagation. The Conditional Probability Neural Network learning algorithm is described as follows:

1 Initialize the weights of the neural network $\theta^{(0)}$;

2 $i \leftarrow 0$;

3 *repeat*

4 $i \leftarrow i + 1$;

5 Calculate

$$\frac{\partial T(\theta)}{\partial \theta} = -\sum_i \sum_j d_{z_i}^{y_j} \left(\frac{\partial b(z_i;\theta)}{\partial \theta} + \frac{\partial f(z_i, y_j;\theta)}{\partial \theta} \right)$$

6 Get $\theta^{(i)}$ by updating $\theta^{(i-1)}$;

7 *Until* $T(\theta^{(i)}) - T(\theta^{(i-1)}) < \epsilon$;

8 $p(y|z;\theta) \leftarrow$ the output of the neural network.

The weight is updated using following algorithm:

1. Perform the forward-propagation phase for an input pattern and calculate the output error.
2. Change all weight values of each weight matrix using the formula:

   Weight (old) + learning rate * output error * output (neurons i) * output (neurons i+1) * (1 - output(neurons i+1))
3. Go to step 1.
4. The algorithm ends, if all output patterns match their target patterns.

IV. Results

A. Dataset

We adopt the FG-NET Aging Database which is publicly available for the experiment. The dataset contains totally 1002 high-resolution color or grey scale face images having variation in lighting, pose and expression. There are 82 subjects in which age ranges from 0 to 69. The real ages are given under the images. A typical aging face images in increasing order in this database is shown in fig.1. The appearance model is used as the feature extractor for the FG-NET database. This model combines the shape and intensity of the face images.

B. Result Set

The performance of the age estimation is measured by Mean Absolute Error (MAE). It is the average absolute difference between the predicted age and the real age. A Leave-One-Person-Out test strategy [6] is used on the FG-NET dataset. i.e., in each fold, facial images of individual person are used as the test set and those of the others are used as the training data. After 82 folds, each subject has been used as test set, and the final results are computed from all of the estimates.

TABLE I
Description Degree of Class Labels

Subject	Labels	Description Degree
001	2	0.0069
	5	0.0174
	8	0.0278
	10	0.0348
	14	0.0487
	16	0.0557
	18	0.0627
	19	0.0662
	22	0.0766
	28	0.0975
	29	0.1010
	33	0.1149
	40	0.1393
	43	0.1498

TABLE II
MAE Result on FG-NET Dataset

	METHOD	MAE
Previous	IIS-LLD	6.27
	CPNN	5.31
Proposed	Modified CPNN	4.48

- **Graphical Result**: Fig.2 shows that the proposed learning algorithm gives less MAE as compared to the existing methods. If MAE of the learning algorithm is lowest then that algorithm is said to be more accurate.

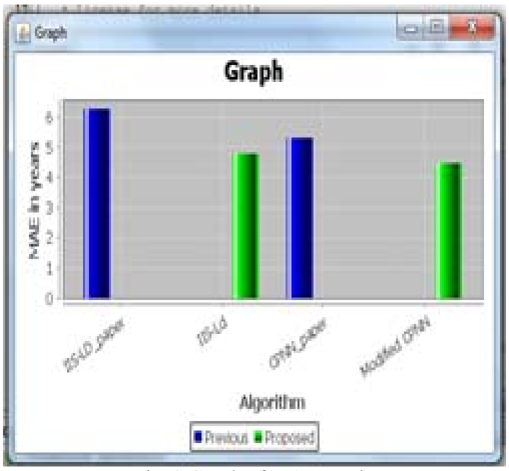

Fig. 2 Graph of MAE result

V. Conclusion

An efficient methods for facial age estimation based on learning from label distribution is proposed in this paper. According to the chronological age of each face image, a label distribution is generated. Then, Label distribution learning algorithms named IIS-LLD and modified CPNN are applied to learn from the generated label distributions. FG-NET aging face database is used to test the learning algorithms. The performance of age estimation methods is measured by mean absolute error. We computed mean absolute error by taking the average absolute difference between the estimated age and chronological age. Graphical results show that the proposed learning algorithm gives less MAE as compared to the existing methods. Finally, we conclude that the proposed learning algorithms perform better than that of previous methods.

References

1. A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation," IEEE Trans. System, Man, and Cybernetics, vol. 34, no.1 pp.621-628, Feb.2004.

2. A. Lanitis, C. Draganova, and C. Christodoulou, "Toward Automatic Simulation of Aging Effects on Face Images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24,no. 4; pp.442-455, Apr 2002

3. G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, " Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," IEEE Trans. Image Processing, vol. 17 no.7, pp. 1178-1188, July 2008.

4. K.-Y. Chang, C.-S. Chen and Y.-P. Hung, "Ordinal Hyper-planes Ranker with Cost Sensitivities for Age Estimation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 585-592, 2011

5. S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning Auto Structured Repressor from Uncertain Nonnegative Labels," Proc. IEEE Int'l Conf. Computer Vision, pp. 1-8, 2007.

6. X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic Age Estimation Based on Facial Aging Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

7. Xin Geng, Chao Yin, and Zhi-Hua Zhou, "Facial Age Estimation by Learning from Label Distributions," IEEE Tran. On Pattern Analysis and Machine Intelligence, vol. 35, No 10.Oct 2013.

8. Y. Fu and T. Huang, "Human Age Estimation with Regression on Discriminative Aging Manifold," IEEE Trans. Multimedia, vol. 10, no. 4, pp. 578-584, June 2008.

9. Y. Fu, Y. Xu, and T.S. Huang, "Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features," Proc. IEEE Int'l Conf. Multimedia and Expo, pp. 1383- 1386, 2007.

10. Y. Zhang and D. Y. Yeung, "Multi-Task Warped Gaussian Process for Personalized Age Estimation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2622-2629, 2010.