Age Estimation based on Support Vector Regression

Junhao Hua *

June 18, 2014

Abstract

While recognition of most facial variations, such as identity, expression,and gender, has been extensively studied, automatic age estimation has rarely been explored. In this paper, we introduce an age estimation technique that combines Active Appearance Models (AAMs) and Support Vector Regression(SVR). The simulations show that SVR model with polynomial kernels got result of $MAE = 6.86$.

Keywords: Age,Estimation, Active Appearance Models, Support Vector Regression

1 Introduction

The determination of the age of a person from a digital photography is an intriguing problem. It involves an understanding of the human aging process, the biomechanical factors that influence the general patterns of aging that the idiosyncratic nature of aging, which is evident in the facial aging differences of identical twins.In this paper,we choose several well-known methods,such as kNN, SVM, and neural networks, to solve an age estimation problem with database of MORPH[4]. The database has 38533 images, and 3004 of them are chosen to be train set to train our feature extracting algorithm. In the later section, we will describe how to solve the age estimation problem in SVR algorithm.

2 Feature Extracting

We use the Active Appearance Models(AAMs)[6, 3] to extract features from image.Firstly, we train the AAM with given images and their coordinate system information. Then we use this model to extracting other images features.After that, we get a database within features(a 1×200 vector) of every images and the corresponding age, gender, Center coordinate(scl),screenshot proportion (trans). We further select age and feature as our database to train SVR algorithm. At last,we get a database with 37814 units. Each unit has an age and feature(a 1×200 vector) of an image.

3 Algorithm Description

3.1 Support Vector Machine (SVM)

We first give a brief overview of the basics of SVMs for binary classification. Then, we explain how this technique can be expanded to deal with the regression problem. Given N training points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ with $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}, i = 1, \ldots, N$ and suppose these points are linearly separable; we have to find a set of N_s support vectors $s_i, (N_s \leq N)$, coefficient weights α_i, constant b_i and the linear decision surface, as in Eq. 1 below, such that the distance to the support vectors is maximized:

$$\mathbf{wx} + \mathbf{b} = 0 \tag{1}$$

*email: huajh7@gmail.com

where

$$\mathbf{w} = \sum_{i=1}^{N_s} \alpha_i y_i s_i. \tag{2}$$

SVMs can be expanded to become nonlinear decision surfaces by first using a mapping Φ to map these points to some other Euclidian space $\mathbb{H}$, that is linearly separable with a given regularization parameter $C > 0, \{\Phi : \mathbb{R}^n \mapsto \mathbb{H}\}$ and by defining a kernel function K, where $K = \Phi(x_i)\Phi(x_j)$. Then, the nonlinear decision surface is defined as:

$$\sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b = 0 \tag{3}$$

where α_i and b_i are the optimal solution of a Quadratic Programming (QP) as follows:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{b},\zeta} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N_s} \zeta_i \\ \textit{subject to} \quad & y_i(\mathbf{w}x_i + mathbfb) \geq 1 - \zeta_i \\ & \zeta_i \geq 0 \end{aligned} \tag{4}$$

3.2 Support Vector Regression(SVR)

The goal of the SVR problem [5] is to build a hyperplane close to as many of the training points as possible. Given N training points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ with $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}, i = 1, \ldots, N$, we have to construct the hyper plane and values of $\mathbf{w}$ and $\mathbf{b}$.In $\epsilon - SV$ regression,our goal is to find a function $f(x)$ that has at most ϵ deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. This is:

$$|y_i(\mathbf{w}x_i + \mathbf{b})|_\epsilon = \begin{cases} 0, & if\ |y_i(\mathbf{w}x_i + \mathbf{b})| \leq \epsilon \\ |y_i(\mathbf{w}x_i + \mathbf{b})| - \epsilon, & otherwise \end{cases} \tag{5}$$

The value of ϵ is selected by the user, and the tradeoff between finding a hyper-plane with a good regression performance is controlled via the given regularization parameter C. The QP problem associated with SVR is described as follows:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{b},\zeta,\zeta^*} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N_s} (\zeta_i + \zeta_i^*) \\ \textit{subject to} \quad & y_i - (\mathbf{w}x_i + \mathbf{b}) \leq \epsilon + \zeta_i \\ & -y_i + (\mathbf{w}x_i + \mathbf{b}) \leq \epsilon + \zeta_i^* \\ & \zeta_i \geq 0, \zeta_i^* \geq 0 \end{aligned} \tag{6}$$

In practice,we use the libsvm* to train our SVR model and select two kernels:Gaussian RBF kernel (Eq.7) and Polynomial kernel (Eq. 8) for comparison with default parameters settings.

$$K(x_i, x_j) = e^{-\frac{1}{2\sigma^2}\|x_i - y_j\|^2} \tag{7}$$

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d \tag{8}$$

*http://www.csie.ntu.edu.tw/ cjlin/libsvm/index.html

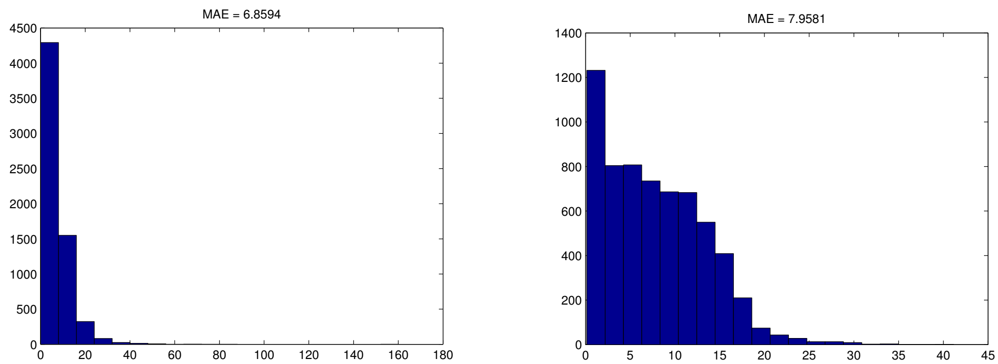

Figure 1: Histogram of absolute age error. left) using polynomial kernel. right) using RBF kernel

4 Experimental Results

This work used a subset of MORPH database[4]. For the performance comparison, we used the Mean Absolute Error (MAE) measurement. MAE is defined as the average of the absolute error between the recognized labels and the ground truth labels:

$$MAE = \frac{\sum_{i=1}^{N_t} |\hat{x}_i - x_i|}{N_t} \tag{9}$$

where $\hat{x}_i$ is the recognized age for the i^{th} testing sample, x_i is the corresponding ground truth, and the N_t is the total number of the testing samples.

The extracted feature dataset is divided into 6 subsets. And we use leave-one-out cross validation for model selection. Namely, we randomly choose 5/6 of dataset for training the SVR model, and 1/6 of dataset for test.

The histogram of absolute age error and the mean-absolute error(MAE) are shown in Fig 1:

From Fig.1 we can see that polynomial kernel has smaller MAE than rbf kernel, which means better accuracy, and most of the estimation bias is less than 10,this is an acceptable result.

MAE is only an indicator of the average performance of the age estimators. It does not provide enough information on how accurate the estimators might be. Suppose there are M test images, $M_{e\leq l}$ is the number of test images on which the age estimation makes an absolute error no higher than l (years), then the *cumulative score* at error level l is calculated by:

$$CumScore(l) = M_{e\leq l}/M \times 100\% \tag{10}$$

If the correct estimation is defined as the estimation with an absolute error no higher than l, then $CumScore(l)$ is actually the accuracy rate. Thus, the cumulative score can be viewed as an indicator of the accuracy of the age estimators. Since the acceptable error level is unlikely to be very high, the cumulative scores at lower error levels are more important.

The cumulative scores of the algorithms and human observers at the error levels from 0 to 15 (years) are compared in Fig.2.

From Fig.2, we can see that SVR with polynomial kernel is overall better than RBF kernel. For polynomial kernel,over 80% of the data tested has error less than 10 years. For RBF kernel, the number is 70%. They are all appropriate kernel.

As we have said,we use leave-one-out cross-validation for train SVR model. Finally, we choose the model with highest accuracy as parameters of the SVR model.

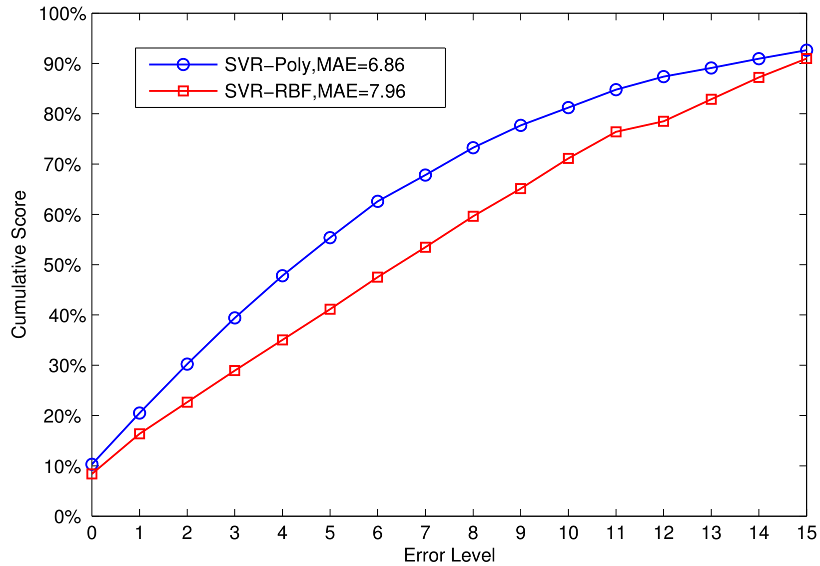

Figure 2: Cumulative scores of standard age estimation on MORPH set

5 Conclusion

In this report, we use SVR model to solve an age estimation problem. The simulations show that SVR model with polynomial kernels got result of $MAE = 6.86$. The code and required data file, please see our attachment.

References

[1] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2234–2240, 2007.

[2] Khoa Luu, Karl Ricanek, Tien D Bui, and Ching Y Suen. Age estimation using active appearance models and support vector machine regression. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, pages 1–5. IEEE, 2009.

[3] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[4] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345. IEEE, 2006.

[5] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[6] Georgios Tzimiropoulos and Maja Pantic. Optimization problems for fast aam fitting in-the-wild. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 593–600. IEEE, 2013.