

AGE REGRESSION FROM FACES USING RANDOM FORESTS

Albert Montillo¹, Haibin Ling²

¹University of Pennsylvania, Radiology; Rutgers University, CIS Dept, Philadelphia, PA USA, montillo@emailtreo.com

²Computer and Information Science Department, Temple University, Philadelphia, PA, USA, hbling@temple.edu

ABSTRACT

Predicting the age of a person through face image analysis holds the potential to drive an extensive array of real world applications from human computer interaction and security to advertising and multimedia. In this paper the first application of the random forest for age regression is proposed. This method offers the advantage of few parameters that are relatively easy to initialize. Our method learns salient anthropometric quantities without a prior model. Significant implications include a dramatic reduction in training time while maintaining high regression accuracy throughout human development.

Index Terms age regression, random forest, learning

1. INTRODUCTION

Human faces contain important attributes which can be used for many computer applications. Attributes such as identity, age, gender and expression form core parameters that can govern applications such as human computer interaction, security and advertising. The attribute of age alone, can be used for many tasks including but not limited to: restricting content to minors and facilitating user requests to retrieve face images from a database.

Automated age estimation would facilitate many applications, however it remains an open research problem due to the challenges involved. Non age-related variations in face images including facial expression, head pose and scale (Fig. 1a-d), confound automated identification of salient age related features. Additionally, age appearance is influenced by both internal factors (true age, genes, gender) and external factors (living location, eating and smoking habits, makeup, and hair dyeing).

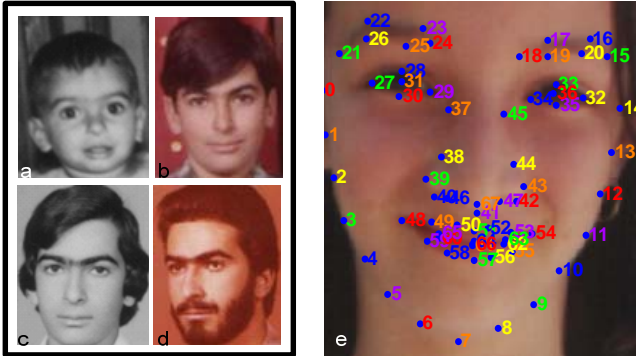


Fig.1 FG-Net (a-d) age variation and (e) landmarks

Age estimation methods can be grouped into three categories [1]: 1) anthropometric methods [4,12] that explicitly model facial shape and texture changes during growth, 2) aging pattern subspace [2] that learns a subspace representation of aging sequences and estimates age by projecting the test face into the subspace, and 3) age regression [1,3,5,6,7,16]; e.g., in [5] a quadratic model for age regression based on least square estimation criterion is constructed. In this paper, we present a method which combines characteristics of both the first and third categories. Age estimation is also related to the problem of age synthesis (e.g., [10]) and face verification (e.g., [8,9]).

Our method performs age regression using information related to anthropometric measurements. However instead of defining a complex cranio-facial model we learn the important anthropometric features automatically. To achieve these goals, we exploit the machine learning method called the random forest [14]. To the best of our knowledge, this is the first use of random forest for age regression of faces.

Renowned for its speed of training and robustness to over-fitting, we hypothesize that the random forest will be a valuable tool for feature selection and prediction in age regression. We derive basic facial measurements from a simple labeling of landmark points around the nose, eyes, chin and mouth. Such measurements are well known to change during development up to the age of 20 years [11], after which face growth plateaus. With a focus on faces aged 0-20 years, our task is to build from a simple labeling both an understanding of which measurements are important and a learner which can perform age regression on novel test faces. In section 2, we briefly review the random forest and the measurements we derive from the landmarks. In section 3, we describe our results on the publicly available FG-Net face aging database [13] (Fig. 1e). In the remaining sections we discuss significance of the rapid speed at which the method estimates feature importance and test error.

2. METHODS

2.1. Problem statement

The problem we set out to address in the paper is: given a training database of face samples consisting of face images, associated face landmarks and a known age of each sample, develop a machine learning method to predict the age on novel untested samples.

2.2. Random forest ensemble learner

The random forest [14] is an ensemble learner consisting of a collection of tree-structured base learners. Each base learner is a classification and regression tree (CART), and for regression, each tree individually predicts the target response while the forest predicts the target as the average of the individual tree predictions.

Formally, a random forest, RF , is a collection of tree-structured learners:

$$RF = \{h_t(\mathbf{X}, \theta_t)\}, t = 1 \dots N_{trees} \quad (1)$$

h_t is the t^{th} individual tree, $h_t(\cdot)$ is the tree's prediction, and N_{trees} is the number of trees. \mathbf{X} denotes our database of samples: $\mathbf{X} = \{S_k\}$ for $k=1 \dots N_{samples}$, where each S_k is a face record (i.e. database sample). \mathbf{X} is divided into two disjoint subsets. One subset is used to train or grow the learner h_t . The other subset is used to test the learner's prediction accuracy. θ_t are i.i.d. random variables. The nature and dimensionality of θ_t depends on the use of randomness in the ensemble construction. For a random forest there are two sources of randomness: 1) The first source of randomness is the selection of a training set from \mathbf{X} known as a bootstrap set, which is used to grow (train and define) each tree. This bootstrap set is a random selection of $N_{samples}$ face records drawn from \mathbf{X} with replacement. 2) The second source is the random selection of m_{try} features (also called predictors) to try as input at each node in a tree being grown. To define a node in a CART-based tree, a decision threshold is selected which partitions the node input samples into two subsets which maximize a purity measure. For a random forest, each tree is grown until nodes have split their inputs into subsets that consist of samples containing just one label (i.e. they are 100% pure).

An important characteristic of the random forest is that while growing each tree, an estimate of test error can be constructed from the training samples. For each tree grown, approximately 36% of the samples in the training set are not selected in the bootstrap and are known as the out of bootstrap (OOB) samples. Using these as input to the corresponding tree enables the random forest to obtain response estimates for them, just as if they were untrained test samples.

2.3. Image landmarks and features

The database consists of the set of face records (database samples) which are 4-tuples containing: a face image, face landmarks, face age and derived features for the face. We denote the database samples as $S_k = \{I_k, L_k, A_k, F_k\}$ where $k=1 \dots N_{samples}$. I_k is the face image, L_k is the set of landmarks on the face I_k , A_k is the age of the face in the image, and F_k is the set of features (i.e., candidate predictors) derived from $\{I_k, L_k\}$ whose nature depends on how we solve the feature selection task. We discuss this feature selection in the next three sections. For each sample, L_k contains the set of N

corresponding landmarks $L_k = \{\ell_i\}$ $i = 1 \dots N$ where ℓ_i are the members of the domain of I_k . For each age we have $A_k \in \mathbb{R}$.

In regression, a challenging task is to define the predictor features, F_k , that facilitate training the base learners so they can accurately predict the target on novel data. A strategy with growing popularity is to define multiple candidate predictors (CP) and then use a learning algorithm to help determine which are most useful. By using a large generous pool of predictors, we increase the chances that we include the useful information for training and learning from the data. In the sections that follow, we adopt such a strategy to define three types of candidate predictors: 1) normalized inter-landmark distances (CPL), 2) ratios of Voronoi cell areas (CPV), and 3) ratios of Delaunay triangle areas (CPT). In summary, for each face sample S_k , we define its derived features F_k as: $F_k = \bigcup (\{CPL_p\}_k, \{CPV_q\}_k, \{CPT_v\}_k)$. Details are explained in the following subsections.

2.4. Normalized inter-landmark distances

We begin with the candidate predictors that are perhaps the easiest to understand. These are the set of inter-landmark distances for each L_k in our database. We denote the set of such distances for L_k as $D_k = \{CPL_p\}$, $p=1 \dots C(N,2)$, where

$$CPL_p = \frac{\|\ell_a - \ell_b\|}{e_k} \text{ where } a, b \in [1 \dots N], b > a, \quad (2)$$

where $\|\cdot\|$ indicates the Euclidean distance, $C(N,2) = N(N-1)/2$ enumerates the number of point pairs for a face sample, and e_k is the Euclidean distance between the eyes for image k , which is used for scale normalization.

2.5. Ratios of Voronoi cell areas

The next type of candidate predictor is composed of the ratio of the areas of Voronoi cells. Given a set L_k of N distinct landmarks in \mathbb{R}^2 , the Voronoi tessellation is the partition of \mathbb{R}^2 into N polygons $vo(\ell_i)$ where $\ell_i \in L_k$. The set of loci in \mathbb{R}^2 that is closer to ℓ_i than any other landmark in L_k is called the Voronoi cell of ℓ_i and is denoted as region $vo(\ell_i)$:

$$vo(\ell_i) = \{x \in \mathbb{R}^2 \mid \|x - \ell_i\| \leq \|x - \ell_j\| \forall \ell_j \in L_k - \ell_i\} \quad (3)$$

We denote the set of cells in the tessellation of L_k as $V_k = \{vo(\ell_i)\}$ where $\ell_i \in L_k$. The tessellation produces one cell per landmark, ℓ_i , hence it is a regular partition of the face sample's image domain into N regions. We define additional candidate predictors for L_k as the ratio of the areas of any two cells of V_k :

$$CPV_q = \frac{area(vo(\ell_a))}{area(vo(\ell_b))} \text{ where } a, b \in [1 \dots N], b > a \quad (4)$$

The number of ratios for every sample is $C(N,2)$. In (4) $area(.)$ is the area of the polygon defined by its Voronoi cell vo argument. By forming ratios for these candidate predictors, we account for variability in the overall scale of the image domain I .

2.6. Ratio of Delaunay triangle areas

Our third type of candidate predictor is composed of the ratio of areas of Delaunay triangles. For a face sample, S_k , given its Voronoi tessellation, V_k , computed in the previous section, we can readily define the Delaunay triangulation of L_k [15]. The tessellation, V_k , consists of polygonal cells whose vertices are called Voronoi vertices. For a given Voronoi vertex, $v \in \mathbb{R}^2$, the nearest neighbor set $nb(L_k, \ell_i)$ of $\ell \in L_k$ consists of the set of points in L_k that are closest to v . The Delaunay triangle for v consists of the triangle formed from the convex hull of the nearest neighbor set of Voronoi vertex v :

$$dt(v) = conv(nb(L_k, v)) \quad (5)$$

We denote the set of triangles in the tessellation of L_k as $T_k = \{dt(v)\}$ where $v \in V_k$. In general, the triangle edges in the Delaunay triangulation will vary somewhat across the face samples in our database due to the variation in landmark location for each L_k . To normalize for such variations, we compute the Delaunay triangulation for one template face sample, S_{k^*} . Hence we preserve the connectivity of the landmarks $\forall k$, however the location, shape and area of the Delaunay triangles vary from one face to the next because the triangle vertices depend on the locations of the landmarks.

We define additional candidate predictors for S_k as the ratio of the areas of any two such distinct Delaunay triangles of a face sample:

$$CPT_v = \frac{area(dt(v_a))}{area(dt(v_b))} \text{ where } v_a, v_b \in V_k, b > a, \quad (6)$$

where $area(.)$ is the area of the specified triangle. The number of ratios for every sample is $C(|V_{k^*}|, 2)$.

3. RESULTS

3.1. Evaluation Methodology

The approach advocated for cross-validation on the FG-Net database is a leave-one-person-out (LOPO) testing strategy [2,16] We adopt this strategy and fold the data along the person dimension. To evaluate our approach, we compute three types of errors: test mean absolute error (test MAE), training mean absolute error (train MAE) and an estimated test error (est. test MAE). Test MAE is the mean absolute difference between the predicted face age and the actual ground truth face age for a sample that is not used to train. Train MAE error is the mean absolute difference between the predicted face age and the actual ground truth face age for a sample that is used to train. The est. test MAE is the

estimate of test MAE using the training samples when those samples are out of bootstrap.

We rank the features used during training by computing, for each candidate predictor, the mean feature importance and then sort them in descending order of importance. We use the feature importance measure based on the increase in node impurity [14].

3.2. Experimental Results Using FG-Net Database

The FG-Net database [13] contains 710 face samples within the age range [0 – 20). These samples are from 81 different subjects. In addition to age, each face sample has been assigned 68 landmarks as shown in Fig. 1e. An example of the computed Voronoi tessellation and Delaunay triangulation for such landmarks is shown for one sample in Fig. 2a and 2b respectively. We use the fast $O(N \log N)$ optimal time *quickhull* algorithm [15] for the computation.

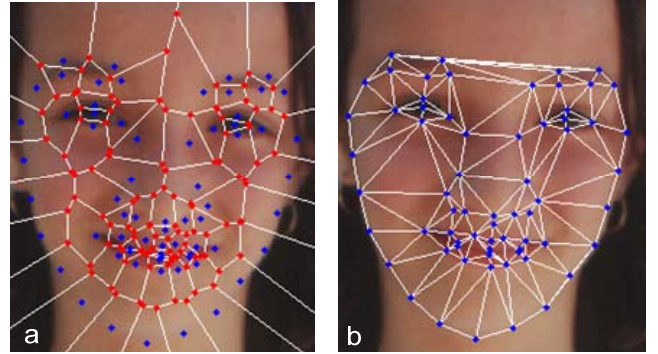


Fig.2 (a) Voronoi cells and their vertices v (red) (b) Delaunay triangles and their vertices ℓ_i (blue landmarks)

Type of candidate predictor	Size	Count
CPL Inter Landmark Distances	$C(68,2)$	2,778
CPV Ratios of Voronoi Cell Areas	$C(68,2)$	2,778
CPT Ratios of Delaunay Triangle Areas	$C(116,2)$	6,670
Total candidate predictor set		11,226

Table1: Candidate predictor (feature) set

Type of error	MAE in years	% variance explained
Test MAE over all folds	3.43	64.34%
Est. test MAE	[3.392 to 3.526]	[62.29% to 65.26%]
Training MAE	0.99	97.01%

Table 2: Test and training errors

Table 1 shows the type and number of candidate predictors used to train the random forest.

Training random forest with the total predictor set described in Table 1 yielded a test MAE of 3.43 years and a training MAE of 0.99 years. In addition, over 81 folds, the est. test MAE had a minimum of 3.392 and maximum of 3.526 years as summarized in Table 2.

The most important predictors for age regression found during our test are those shown in Fig 3.

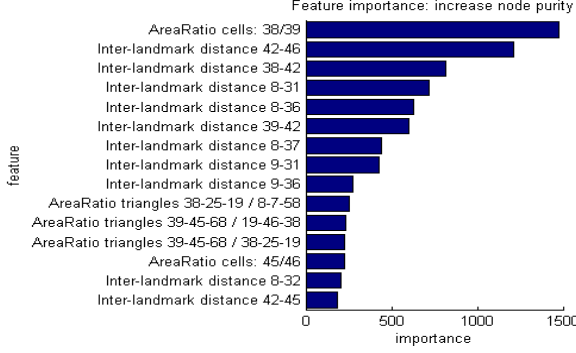


Fig.3 Feature importance from increase in node purity

Fig 4a shows that test MAE drops rapidly and stays low during training, indicating we have not overfit the learner.

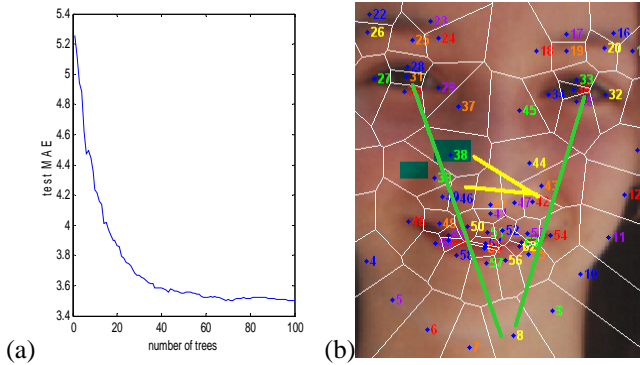


Fig.4: (a) Test MAE as a function of # trees. (b) Location of most important features

4. DISCUSSION

Achieving age regression accuracy on test data of just 3.4 years is a promising sign for further experimentation. Research has shown that humans are able to predict age to within 3.6 years on the full database of 1002 samples [5], and our method performs close to humans. Also our testing to date does not exploit appearance information, which is expected to bring further improvement.

One of the most potentially valuable results is that estimated test error from OOB training samples is quite reliable. The range of the estimated test MAE was 3.392 to 3.526 over 81 folds. Hence the maximum deviation of the estimated test MAE on any given single fold from the actual test MAE computed over ALL folds is only 0.09 years. This means that by using the random forest, we obtain an accurate prediction of test error in one fold. This can result in a reduction of training time by a factor of $N_{fold}-1$ where N_{fold} is the number of cross validation folds. For large databases the savings can be substantial. For example, in the FG-Net database, 81 folds are performed for LOPO folding requiring ~8hrs on a 3GHz Pentium, while one fold requires only ~10min, a substantial reduction in time.

The predictors that random forest ranks as most important appear to be the ones closely related to the growth of the face during human development. A few of the top

ranking predictors are illustrated in Fig 4b. These mostly have to do with the elongating of the nose and the overall height of the face relative to eye separation which changes much more slowly. The advantage of the random forest is that it is fast to train, and gives insight into which features help predict the target.

5. CONCLUSIONS

This paper presents the application of the random forest for face age regression. The promising experimental results for subjects aged 0-20 years, show the effectiveness of the proposed method. Furthermore, RF has only two adjustable parameters (N_{trees} and m_{try}) and we found that the results are insensitive to them. Another interesting observation is that the estimated test error (obtained during training) very well approximates the actual test error. In the future we plan to investigate (1) the combination of geometric and appearance features and (2) the extension to age estimation over all age ranges.

References

- [1] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," ICME, 2007.
- [2] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," IEEE Trans. PAMI, 29(12):2234–2240, 2007.
- [3] G. Guo, Y. Fu, C. R. Dyer, T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," IEEE Trans. Image Proc., 17(7):1178–1188, 2008.
- [4] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," CVIU, 74(1):1–21, 1999.
- [5] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," IEEE Trans. Syst Man Cybern B Cybern., 34(1):621–628, 2004.
- [6] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Trans. PAMI, 24(4): 442–455, 2002.
- [7] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," ICCV, 2005.
- [8] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "A study of face recognition as people age," ICCV, 2007.
- [9] N. Ramanathan and R. Chellappa, "Face verification across age progression," IEEE Trans. Image Proc., 15(11):3349–3361, 2006.
- [10] J. Suo, F. Min, S. Zhu, S. Shan, X. Chen, "A multi-resolution dynamic model for face aging simulation," CVPR, 2007.
- [11] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception". J. of Exp. Psychology: Human Perception and Performance, 1(4):374-382, 1975.
- [12] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," CVPR, 2006.
- [13] The FG-NET Aging Database, <http://www.fgnet.rsunit.com/>.
- [14] L. Breiman, "Random forests," Machine Learning, 45:5-32, 2001.
- [15] C. Barber, D. Dobkin, & H. Huhdanpaa, "Quickhull algorithm for convex hulls," ACM Trans. Math. Softw. 22(4):469-483, 1996.
- [16] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," ICCV, 2007.