Exploiting Unlabeled Ages for Aging Pattern Analysis on A Large Database

Chao Zhang and Guodong Guo
Lane Dept. of CSEE, West Virginia University, Morgantown, WV 26506
cazhang@mix.wvu.edu, guodong.guo@mail.wvu.edu

Abstract

"Big Data" analysis is an emerging topic in computer vision and pattern recognition. As one example problem of big data, we study semantic age labels and facial aging pattern analysis on a large database. In aging analysis, one of the great challenges is the lack of a large number of face images with ground truth age labels. Unlike many other example-based recognition problems where human annotations can be used as the ground truth labels for both training and testing, it is quite difficult to label the exact ages in face images by human annotators. An alternative is to exploit the unlabeled ages to enhance the age estimation performance. However, it is unclear whether the face images with unlabeled ages can be used or not for age estimation, and how to use the unlabeled data. In this paper, we study the two problems comprehensively under two paradigms: the semi-supervised learning and unsupervised learning for aging pattern analysis. We emphasize the importance of using ground truth age labels and a large database in order to derive a meaningful measure in the context of big data. Our study can make an impact on collecting aging patterns that is very expensive and time consuming in practice.

1. Introduction

"Big Data" analysis is an emerging topic in computer vision and pattern recognition. As one example problem of big data, we study semantic age labels and facial aging pattern analysis on a large database. Human age estimation has a number of potential applications, including business intelligence [19] and age-specific human-computer interaction [7]. The useful applications have attracted researchers to study age estimation deeply and broadly [6] [17], in order to make age estimation more robust and reliable for practical use.

There are many challenges to develop a good age estimator. In reality, different people may age quite differently, because of the gene difference as well as life style and living condition differences. In age estimation research, probably the most difficult thing is the aging database collection. It is much more difficult than collecting a database for face recognition. A large span of ages, e.g., from very young to much older, need to be collected with ground truth ages.

In aging database collection, it is important to have the true age label for each face image, since it is not trivial to annotate the age by humans for each face image after the database collection. In a very recent study [28], the authors have shown that perceiving age in face photos by humans is a challenging task. The mean absolute error (MAE) for human age perception is 8.58 years on selected face images from MORPH [18], and is 8.13 years on face images from FG-NET [5]. These reported errors on human age perception are even much higher than some computational age estimation methods, e.g., [9] [14] [29] [15] [10] [22], with MAEs about or below 5 years on the same databases.

It has also been shown in [16] that the face aging database collected from the Internet contains outliers of age labels. Even a robust age estimator was developed to deal with the age label outliers, the age estimation results still have large errors. For example, it was reported that when the age estimator is learned from the collected face database (with incorrect age labels), and then applied to some standard databases (with ground truth age labels), the MAE could be 8.60 years on MORPH [18], and 9.49 years on FG-NET [5].

So, age labeling is a big issue in the context of "big data." It is very important to have age labels for both training and testing. The performance measure (e.g., MAE) might be incorrect or inaccurate when the ground truth age labels are not available for the test face images. Basically, we cannot rely on human annotations of the age labels for face images, either training or testing. This is very different from some other computer vision problems, such as the general object recognition [27], action recognition [13], and facial attributes [12]. In those problems, the required labels can be obtained from human annotators, e.g., Amazon's Mechanical Turk, without much difficulty.

Is there a way to deal with the age labeling problem? Suppose there is a collected face database which contains a number of face images *without* age labels. Should we only use the face images with ground truth age labels? Or can

we utilize the *unlabeled* ages (i.e., face images without age labels) to improve the age estimation performance?

In this paper, we study the age labeling problem in age estimation. We explore the feasibility of using face images with unlabeled ages to improve the age estimation performance. An illustration of the problem is shown in Figure 1. Towards this goal, we put the age labeling problem into two paradigms: semi-supervised learning and unsupervised clustering. Both paradigms involve unlabeled ages, in contrast to the traditional supervised age estimation [6] [17]. We measure the age estimation performance under these two paradigms, and compare with the supervised learning based estimation to assess the usefulness of unlabeled face images. If the age estimation performance is sufficiently good using a small amount of labeled ages in combination with a number of unlabeled face images, there may be no need to turn to age labeling by human annotations. Human annotations of age labels are not accurate, and incorrect age annotations might hurt the age estimation accuracy [16].

Further, we study the problem of how many labeled examples are needed for age estimation to have a reasonably good performance on a large database. This is valuable to reduce the cost and time for aging database construction.

In our study, we emphasize the importance of ground truth age labels for a meaningful and quantitative measure. Otherwise, the conclusions may not make any sense as a practical guide to collecting aging patterns and age labels.

We notice that there was a semi-supervised approach to perceived age prediction [21], but their goal is very different from ours. First, there is no ground truth age labels for the research in [21], so it is difficult to tell if there is any advantage to use unlabeled ages for age estimation; Second, the work in [21] is about perceived age, while we focus on chronological age estimation by a computer; Third, we found that the clustering-based annotation of perceived ages cannot work well in our experiment with ground truth ages (see Section 4), since the clustering cannot find correct cluster centers for the complex and subtle aging patterns.

Our *major contributions* in this paper include: (1) quantitatively assessing the usefulness of unlabeled ages for aging pattern analysis under different paradigms; and (2) discovering the required amount of labeled ages to develop a robust estimator through the combination of unlabeled ages.

In the remaining of the paper, we exploit the unlabeled ages under two paradigms in Sections 2 and 3, respectively. Several learning methods are investigated within each paradigm. Then we present experiments on a large database in Section 4. Finally, we draw conclusions.

2. Exploiting Unlabeled Ages by Semi-Supervised Learning

To assess the usefulness of unlabeled ages for aging pattern analysis, we place the problem into two learning
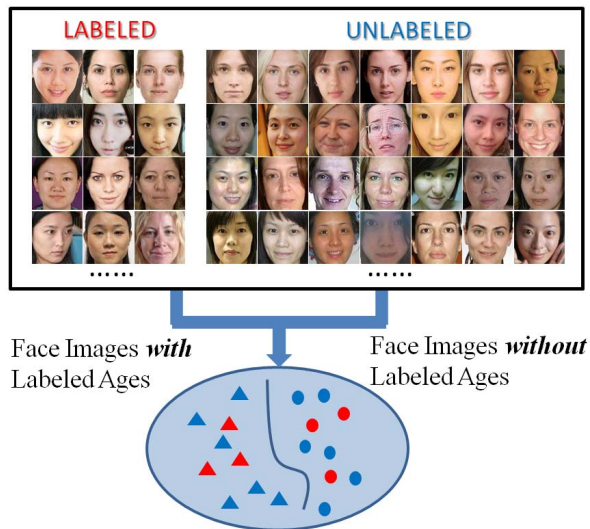

Figure 1. Illustration of the problem we study: Is it feasible to utilize the unlabeled ages for aging pattern analysis? And how to use them? (The red patterns are labeled ages, and blue unlabeled.)

paradigms: (1) semi-supervised learning [30], and (2) unsupervised learning [25]. We explore the semi-supervised learning paradigm in this section, and unsupervised learning next. Both paradigms are compared to a supervised learning method to have a quantitative measure. The supervised learning is based on the support vector machines (SVMs) [23] trained on the labeled ages only.

We investigate the feasibility of executing semi-supervised learning for aging pattern analysis, and discover how useful the unlabeled ages could be for age estimation. We also study the required amount of labeled ages to develop a robust age estimator.

There are a number of techniques developed for semi-supervised learning [30]. For our study, we investigate three methods to exploit the unlabeled ages. A brief description of the methods is presented below.

2.1. Semi-supervised Discriminant Analysis

Suppose we have a set of l face images with age labels $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, and a set of u face images without age labels $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$. The classic Linear Discriminative Analysis (LDA) aims to find a projection vector $\mathbf{a}$ such that the ratio of $\mathbf{a}^T S_b \mathbf{a}$ over $\mathbf{a}^T S_t \mathbf{a}$ is maximized, where the between-class scatter matrix S_b and total scatter matrix S_t (the summation of the within-class and between-class scatter matrices) are computed from the labeled data set.

A regularizer can be imposed to the standard LDA to change the optimization problem into the following [11],

$$\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_t \mathbf{a} + \alpha J(\mathbf{a})} \qquad (1)$$

where the regularizer $J(\mathbf{a})$ controls the learning complex-

ity and the coefficients α balances between the model complexity and the empirical loss. The regularizer $J(\mathbf{a})$ also provides the flexibility to incorporate the manifold structure of the unlabeled examples when they are available [2].

The unlabeled data are used to construct a p-nearest neighbor graph G to model the relationship between nearby data points. Thus, a natural regularizer can be defined as follows:

$$J(\mathbf{a}) = \sum_{ij}(\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j)^2 W_{ij} = 2\mathbf{a}^T X L X^T \mathbf{a} \quad (2)$$

$L = D - W$ is the Laplacian matrix. W is the weight matrix defined by

$$W_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_p(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_p(\mathbf{x}_i) \\ 0, & \text{otherwise}, \end{cases} \quad (3)$$

where $N_p(\mathbf{x}_i)$ denotes the set of p nearest neighbors of $\mathbf{x}_i$. D is a diagonal matrix; its entries are the column sum of W, $D_{ii} = \sum_j W_{ij}$.

Using the regularizer derived from unlabeled face images, the objective function of semi-supervised discriminant analysis (SDA) [2] is given by,

$$\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T(S_t + \alpha X L X^T)\mathbf{a}}. \quad (4)$$

The projection vector $\mathbf{a}$ that maximizes the objective function is given by the solution of the following generalized eigenvalue problem:

$$S_b\mathbf{a} = \lambda(S_t + \alpha X L X^T)\mathbf{a} \quad (5)$$

The solution of (5) is a set of eigenvectors learned from face images with and without age labels for aging pattern analysis.

2.2. Manifold Regularization

Belkin et al. [1] proposed a graph-based manifold regularization (MR) framework that allows to incorporate the intrinsic geometrical structure of the input data as a regularization term. It can be performed as semi-supervised learning when unlabeled samples are available. The manifold regularization framework employs two regularization terms:

$$f^\star = \arg\min_f \frac{1}{l} \sum_{i=1}^{l}(f(\mathbf{x}_i) - y_i)^2 + \gamma_A\|f\|_K^2 + \gamma_I\|f\|_I^2 \quad (6)$$

where the first term is an arbitrary loss function, in our case, it is defined as the residue square error over all labeled examples. K is a 'base kernel', e.g., a linear or RBF kernel. I is a regularization term induced by the labeled and unlabeled data. γ_A controls the complexity of the function in the ambient space, while γ_I controls the complexity of the function in the intrinsic geometry.

Sindhwani et al. [20] gave a semi-supervised kernel that can cover all input data, including both labeled and unlabeled. The kernel supports induction. Then the objective function becomes:

$$f^\star = \arg\min_f \frac{1}{l} \sum_{i=1}^{l}(f(\mathbf{x}_i) - y_i)^2 + \gamma_A\|f\|_K^2 +$$
$$\frac{\gamma_I}{(u+l)^2} \sum_{i,j=1}^{l+u}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 W_{ij} \quad (7)$$

where W_{ij} is defined by Eq. (3).

According to [1], the minimizer of problem (7) admits an expression

$$f^*(\mathbf{x}) = \sum_{i=1}^{l+u}\alpha_i k(\mathbf{x}, \mathbf{x}_i) \quad (8)$$

where $\alpha = (\alpha_1, \ldots, \alpha_{l+u})^T$ are the parameters to be learned by either a least square solver or a SVM quadratic programming solver [1], T denotes the transpose, and $k(\mathbf{x}, \mathbf{x}')$ is a kernel function. l and u denote the number of labeled and unlabeled face images, respectively.

Expression (8) is used to map an aging pattern $\mathbf{x}$ to the corresponding age label.

2.3. Spectral Regression Discriminant Analysis

Let $\bar{\mathbf{x}}_i = \mathbf{x}_i - \mu$ be the centered aging pattern, where μ is the total mean vector over the training set. $\bar{X}^{(k)} = [\bar{\mathbf{x}}_1^{(k)}, \cdots, \bar{\mathbf{x}}_{m_k}^{(k)}]$ is the centered data matrix of kth age class, and m_k is the number of aging patterns in the kth class. Then the between class scatter matrix can be re-written as

$$S_b = \sum_{k=1}^{c}\bar{X}^{(k)}W^{(k)}(\bar{X}^k)^T = \bar{X}W\bar{X}^T, \quad (9)$$

and the total scatter matrix is $S_t = \bar{X}\bar{X}^T$, where $\bar{X}$ is the centered data matrix, and $W^{(k)}$ is a $m_k \times m_k$ matrix with all the elements equal to $1/m_k$. W is a $m \times m$ block-diagonal matrix, assuming there are m training examples,

$$W = \begin{bmatrix} W^{(1)} & 0 & \ldots & 0 \\ 0 & W^{(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & W^{(c)} \end{bmatrix}.$$

Therefore the generalized eigenproblem for LDA can be represented as

$$\bar{X}W\bar{X}^T\mathbf{a} = \lambda\bar{X}\bar{X}^T\mathbf{a}, \quad (10)$$

Let $\bar{y}$ be the eigenvector of the eigenproblem

$$W\bar{y} = \lambda\bar{y} \qquad (11)$$

with eigenvalue λ, and let $\bar{X}^T\mathbf{a} = \bar{y}$, then $\mathbf{a}$ is the eigenvector of the eigenproblem in (10) with the same eigenvalue λ, according to [3]. Thus, instead of solving the eigenproblem in (10), the LDA basis functions can be obtained through two steps: (1) solve the eigenproblem (11) to get $\bar{y}$; and (2) find the $\mathbf{a}$ that satisfies $\bar{X}^T\mathbf{a} = \bar{y}$ in the least squares sense, $\mathbf{a} = \arg\min_{\mathbf{a}} \sum_{i=1}^{m}(\mathbf{a}^T\bar{\mathbf{x}}_i - \bar{y}_i)^2$, where $\bar{y}_i$ is the ith element of $\bar{y}$. A regularization term is usually used to derive a stable solution for the least squares regression,

$$\mathbf{a} = \arg\min_{\mathbf{a}} \left(\sum_{i=1}^{m}(\mathbf{a}^T\bar{\mathbf{x}}_i - \bar{y}_i)^2 + \alpha\|\mathbf{a}\|^2 \right). \qquad (12)$$

Because the above two-step procedure involves spectral analysis of the graph matrix W and the regression technique, it is called spectral regression discriminant analysis (SRDA) [3]. It can be generalized to deal with semi-supervised age estimation by constructing the graph matrix W using both labeled and unlabeled aging patterns.

3. Unsupervised Learning of Aging Patterns

To further assess the usefulness of unlabeled ages, we also study the unsupervised learning paradigm for aging pattern analysis, in addition to the semi-supervised paradigm. For unsupervised aging pattern analysis, we use the classic k-means clustering algorithm and a more advanced spectral learning technique [24].

To test the clustering result of the aging patterns for each algorithm, we compare the clusters with the ground truth age labels. Specifically, the majority of the age labels within each cluster is used as the cluster's age label, and assigned to each pattern within that cluster as the "estimated" age. Then we can compute the estimation errors for the whole data set.

3.1. K-Means Clustering of Aging Patterns

We use the k-means clustering method for unsupervised aging pattern analysis. The classic k-means algorithm remains a popular choice for real-world clustering tasks [26]. We explore if the aging patterns can be clustered into different age classes using the k-means method.

The k-means optimization problem is to find k cluster centers $\mathbf{c} \in \mathrm{R}^m$ from a set of aging patterns $\mathbf{x}_i \in \mathrm{R}^m$ by minimizing the following objective function: $min \sum_{\mathbf{x}_i \in X} \|f(C, \mathbf{x}_i) - \mathbf{x}_i\|^2$, where $f(C, \mathbf{x}_i)$ returns the nearest cluster center $\mathbf{c} \in C$ to $\mathbf{x}_i$ from all clusters C using Euclidean distance. The optimization process starts with k initial seeds (randomly selected) and iterates until convergence to a local optimum. We attempt different k values to measure the difference.

3.2. Spectral Clustering of Aging Patterns

Spectral clustering [24] is a category of more advanced unsupervised learning techniques, which are usually more effective to find clusters and can outperform some traditional methods. Here we exploit if there is any advantage to use spectral clustering for aging pattern analysis. The basic idea of spectral clustering is to measure the similarity between pairs of aging patterns and construct a similarity graph $G = (V, E)$, where each vertex v_i represents one aging pattern $\mathbf{x}_i$. Two vertices are connected if the similarity s_{ij} between the corresponding patterns $\mathbf{x}_i$ and $\mathbf{x}_j$ is positive or larger than a certain threshold, and the edge is weighted by s_{ij}. A typical similarity function is computed by $s_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$, where σ is a scaling parameter to adjust the similarity measure. The clustering problem is then to partition the similarity graph into groups such that the edges between different groups have very low weights and the edges within a group have high weights.

The spectral clustering can have various similarity measures and be solved with different methods. In our aging pattern analysis, we adopt a very recent technique [4] to cluster the aging patterns.

4. Experiments

Now we exploit the unlabeled ages for aging pattern analysis based on several experiments. First, the database is introduced for this study. Second, the experimental designs are described. Third, the experimental results based on the semi-supervised learning paradigm are presented. Fourth, the results of unsupervised learning or clustering are given. Finally, a comparison with an exiting approach is performed on our database with ground truth age labels.

4.1. Database and Aging Pattern Representation

We use the MORPH-II [18] database for our study, which contains more than 50,000 face images of different gender and ethnicity along with a span of ages from 16 to 67 years. All face images have the ground truth labels of age, gender, and ethnicity, which are important for our quantitative study of the age labeling problem. In MORPH-II, however, the distribution of gender and ethnicity is very unbalanced. For example, it has about 77% Black faces, 19% White, and 4% other races, e.g., Hispanic, Asian, and Indian. There are also more males than females. To study the age labeling problem on a relatively balanced distribution of various populations, we selected partial data from MORPH-II, following the suggestion in previous research [8]. Almost all White face images are used, and a matched number of Black faces is used to balance the number of face images in the two races. In total, we have a database of 21,060 face images chosen from the MORPH. In the selected database, there are 2,570 White Female (WF), 7,960

White Male (WM), 2,570 Black Female (BF), and 7,960 Black Male (BM) face images. To avoid the influence of gender and ethnicity on our studies, our age estimations are performed on each gender and ethnicity group separately. Then the average over the four groups is computed as the final quantitative measure.

The face images in the database are aligned with detected eye centers, and cropped and normalized into a fixed size, e.g., 60 x 60. For the aging pattern representation in face images, we use one of the state-of-the-art methods to extract the biologically inspired features (BIF) [10]. The support vector machines (SVMs) [23] are used for learning the aging functions in each case of age estimation.

4.2. Experimental Design

To exploit the unlabeled ages for aging pattern analysis, we design age estimation experiments under two different paradigms: (1) semi-supervised learning, and (2) unsupervised learning or clustering. The results are compared to the traditional supervised learning based age estimation, which is considered as the baseline. In supervised learning, only the labeled ages are used to learn the aging functions with the SVMs, while the unlabeled ages are not used. The aging patterns for testing are the same for all methods.

Under either paradigm, we perform 10 random runs of the methods and compute the average and standard deviation. The purpose is to get a statistically meaningful results and avoid possible perturbations caused by the randomness.

Under the semi-supervised learning paradigm, we also study the age estimation performance with different percentages of labeled and unlabeled ages, so that we can measure how many labeled ages are needed to develop a robust age estimator. This quantitative measure is of great value in practice, considering the expensive and time-consuming aging pattern collection. Three methods are investigated under the semi-supervised learning paradigm to compare with each other and with the supervised learning results. All three methods are new to aging pattern analysis.

Under the unsupervised learning or clustering paradigm, we will discover if the clustering based methods can work or not, and evaluate the results using different numbers of cluster centers. The classic k-means clustering algorithm and an advanced spectral clustering method are used for this investigation.

From our data set of 21,060 face images, 2,112 is randomly selected as the fixed test set, which is about 10% of the whole data set. Then the remaining 18,948 face images are used as the training set. In the training set, we randomly selected a different amount of data, e.g., 10%, 20%, ···, 90%, as the labeled ages, while the remaining as the unlabeled ages in which the age labels are assumed unknown for the semi-supervised learning paradigm.

4.3. Results under Semi-Supervised Learning

The age estimation results under the semi-supervised learning paradigm are shown partially in Table 1. The age estimation is performed within each gender and ethnicity group, and the average over the four groups is shown in Figure 2. The MAEs of age estimation are measured w.r.t. the different amount of labeled ages, ranging from 10% to 90%. We can observe that (1) when more labeled ages are used for training, the age estimation errors are reduced for each semi-supervised method or the supervised learning approach; (2) the investigated semi-supervised methods are consistently better than the supervised learning approach, which demonstrates that the unlabeled ages are useful to improve the age estimation performance compared to the supervised learning, using the same amount of labeled ages. Thus the unlabeled ages can be helpful under the semi-supervised learning paradigm; (3) the three semi-supervised learning methods perform differently in age estimation: the SDA is better than the MR, and the SRDA performs the best among the three; (4) the age estimation performance is different in the four gender and ethnicity groups, the White Male group has the smallest MAEs, the White Female and Black Male have comparable MAEs, and the Black Female has the highest MAEs; and (5) about 20-30% of labeled ages can make the MAE reduced to about one year difference from the 90% labeled ages. For a better method, e.g., SRDA, the MAE difference (from the 90% labeled ages) is smaller than other methods.

The average MAEs over the four groups are also shown in Table 1 partially, ranging from 10% to 50% and the last one 90% of labeled ages. The MAEs corresponding to different percentages of labeled ages are compared to the 90%, and the MAE differences are computed and shown in the table. We can see that about 20-30% of labeled ages can be good enough to reduce the MAE to about one year difference from the 90% labeled ages. This finding is very useful and valuable in practice. A small percentage of age labels can be collected together with a large number of unlabeled ages to develop a good age estimator. This new finding has not been discovered by any previous research, to the best of our knowledge.

4.4. Results under Unsupervised Clustering

The results of exploiting unlabeled ages under the unsupervised clustering paradigm are shown in Figure 3. Only the aging patterns for testing are used for this experiment, since there is no training data needed in this paradigm. As stated in Section 3, the majority of the patterns within each cluster is considered as the cluster's age label. Then the ground truth age labels can be used to compare with the cluster labels for MAE measure. The averages over ten random runs with the deviations are shown for each clustering result. The clustering is performed on the four gender and

Table 1. Exploiting the unlabeled ages with the semi-supervised learning paradigm, comparing to a supervised learning method where only the labeled ages are used for learning. The age estimation results are measured by the popularly used mean absolute error (MAE) in years. Using 90% labeled ages gives better results than less percentages in each case, and it is used as the basis to measure the MAE difference (denoted by "Diff.") when other percentages of labeled ages are used. "MR" represents the manifold regulation method, "SDA" for semi-supervised discriminant analysis, and "SRDA" for spectral regression discriminant analysis.

Method	Percentage of Labeled Ages (%)										
	10%		20%		30%		40%		50%		90%
	MAE	Diff.	MAE	Diff.	MAE	Diff.	MAE	Diff.	MAE	Diff.	MAE
Supervised	7.52	1.97	6.94	1.39	6.67	1.12	6.39	0.84	6.17	0.62	5.55
MR	7.07	2.44	6.34	1.71	5.84	1.21	5.62	0.99	5.25	0.62	4.63
SDA	6.34	2.00	5.48	1.14	5.13	0.79	4.83	0.49	4.64	0.30	4.34
SRDA	5.73	1.67	5.07	1.01	4.79	0.73	4.58	0.52	4.45	0.39	**4.06**

ethnicity groups and then averaged. Different numbers of cluster centers are attempted, ranging from 50 to 300, for both the k-means and spectral clustering methods. We find that the clustering methods cannot work well for aging pattern analysis. The advanced spectral clustering approach actually results in higher errors than the classic k-means clustering algorithm. This observation is different from those on many databases in machine learning, revealing the special property of the aging patterns. Both clustering methods have much higher errors compared to the semi-supervised or supervised learning results shown in Table 1 or Figure 2. As a result, the unsupervised clustering paradigm cannot work well for unlabeled aging pattern analysis. The reason may be that the aging patterns are subtle and could not be grouped naturally into separate clusters.

Figure 3. The results of aging pattern analysis under the unsupervised clustering paradigm. The MAEs are averaged over ten random runs and the deviation bars are shown. The results are measured with different numbers of cluster centers, from 50 to 300. The advanced spectral clustering method performs even worse than the classic k-means clustering algorithm, and both cannot perform well for aging pattern analysis. This empirical study shows that aging patterns are too complex to do unsupervised clustering.
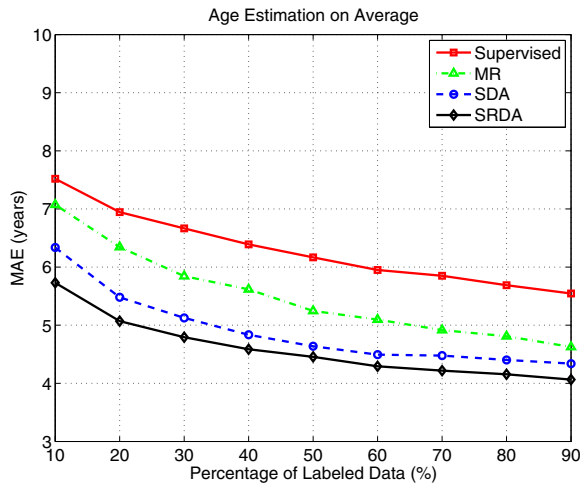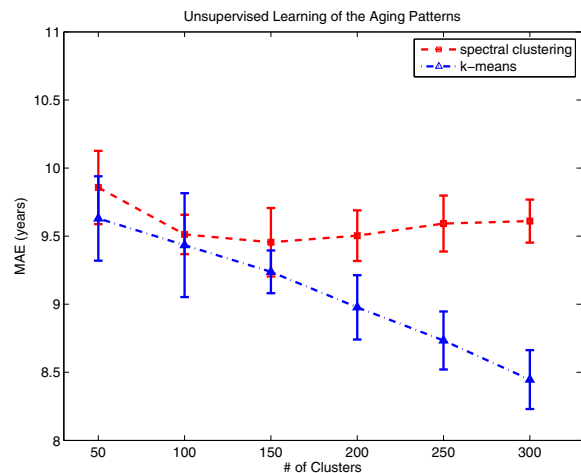

Figure 2. The average performance of age estimation over the four groups of gender and ethnicity. The three semi-supervised learning based age estimation methods (MR, SDA, and SRDA) are consistently better than the supervised learning method.

4.5. Comparison with Another Approach

We noticed that there was a semi-supervised approach to perceived age prediction [21], where there are no ground truth age labels in their experiment. The study in [21] cannot tell whether the unlabeled ages are useful or not for aging pattern analysis. Although the goal of [21] is different from ours, we plan to compare the scheme in [21] with our approaches to discover more things. We implemented the method in [21] by ourselves and do a similar age prediction experiment on our database where all ground truth age labels are available. The result based on the method in [21] is shown in the first row in Table 2. The MAE is 9.70 years, which is significantly worse than our approaches. It is about 91.3% higher than our result based on the SRDA method. The results of our approaches are shown in rows 2-4 in Table 2, where 20% labeled ages are used for our semi-supervised learning, which is the minimum amount needed based on our quantitative study shown in Table 1 and Figure 2, respectively.

Table 2. Age estimation using the semi-supervised approach in [21] based on our implementation and evaluated on our database. The emphasis of only 200 labeled ages [21] and other reasons result in a MAE of 9.70 years, which is much higher than our approaches to age estimation based on semi-supervised learning methods using 20% of labeled ages for training.

Method	#Labeled	#Unlabeled	# Test	MAE
[21]	200	18748	2112	9.70
MR(ours)	3800	15148	2112	6.34
SDA(ours)	3800	15148	2112	5.48
SRDA(ours)	3800	15148	2112	5.07

In [21], the emphasis is their active learning scheme to determine the number of age labels to be annotated by humans, which is actually a k-means clustering of the unlabeled data. Only 200 cluster centers are used in [21] based on their active learning on their large database comparable to the size of our database. Based on our investigation of the unsupervised learning paradigm in Section 4.4, the clustering method is not proper for aging pattern analysis.

In addition to the problem caused by the clustering, there are some other issues that interpret the high errors using the approach in [21]. One is that the number of labeled examples, 200, advocated in [21] is too small to provide sufficient labeled ages, another is that the semi-supervised learning method used in [21] might not be so good as the methods we explore here, and there is no consideration of gender and/or ethnicity groups in their perceived age prediction [21].

5. Conclusions

We have studied how to exploit the unlabeled ages for aging pattern analysis using two paradigms, semi-supervised learning and unsupervised clustering. Several methods have been explored under the two paradigms for our quantitative study. Based on our empirical evaluations on a large database, we have discovered that the unlabeled ages are useful to enhance the age estimation performance when they are properly utilized. The semi-supervised learning paradigm is appropriate to take advantage of the unlabeled ages, while the unsupervised clustering paradigm cannot work well for aging pattern analysis. The reason might be that the aging patterns are very subtle and their distribution is too complex to have a natural separation by clustering. The labeled ages are necessary in exploiting unlabeled ages. Based on our quantitative study, we found that about 20% to 30% of labeled ages can be sufficient to develop a robust age estimator, comparing to using almost all labeled ages (e.g., 90%). This finding can make an impact on collecting aging patterns and age labels, which are very expensive and time consuming in the context of "big data."

References

[1] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.

[2] D. Cai, X. He, and J. Han. Semi-supervised discriminant analysis. In *IEEE 11th Int'l Conf. on Computer Vision*, pages 1–7, 2007.

[3] D. Cai, X. He, and J. Han. Srda: An efficient algorithm for large-scale discriminant analysis. *IEEE Trans. on Knowledge and Data Engineering*, 20(1):1–12, 2008.

[4] W. Chen, Y. Song, H. Bai, C. Lin, and E. Chang. Parallel spectral clustering in distributed systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(3):568–586, 2011.

[5] FGNET. The fg-net aging database. In *http://www.fgnet.rsunit.com/*, 2002.

[6] Y. Fu, G.-D. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(11):1955–1976, 2010.

[7] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. on PAMI*, 29(12):2234–2240, 2007.

[8] G.-D. Guo and G. Mu. Human age estimation: what is the influence across race and gender? In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2010.

[9] G.-D. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *IEEE Conf. on CVPR*, pages 657–664, 2011.

[10] G.-D. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119, 2009.

[11] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[12] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *IEEE International Conference on Computer Vision*, Oct 2009.

[13] I. Laptev, M. Marszaek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *CVPR*, 2008.

[14] C. Li, Q. Liu, J. Liu, and H. Lu. Learning ordinal discriminative features for age estimation. In *IEEE CVPR*, pages 2570–2577, 2012.

[15] K. Mitra, A. Veeraraghavan, and R. Chellappa. Robust rvm regression using sparse outlier model. In *CVPR*, pages 1887–1894, 2010.

[16] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *ACM Multimedia*, 2009.

[17] N. Ramanathan, R. Chellappa, and S. Biswas. Age progression in human faces : A survey. *Visual Languages and Computing*, 2009.

[18] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE Conf. on AFGR*, pages 341–345, 2006.

[19] C. Shan, F. Porikli, T. Xiang, and S. Gong. *Video Analytics for Business Intelligence*, volume 409. Springer, 2012.

[20] V. Sindhwani, P. Niyogi, and M. Belkin. Beyond the point cloud: from transductive to semi-supervised learning. In *Proc. of the 22nd int'l conf. on Machine learning*, pages 824–831, 2005.

[21] M. Sugiyama and Y. Ihara. A semi-supervised approach to perceived age prediction from face images. *IEICE Trans. on Information and Systems*, 93(10):2875–2878, 2010.

[22] J. Suo, S. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):385–401, 2010.

[23] V. N. Vapnik. *Statistical Learning Theory*. John Wiley, New York, 1998.

[24] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[25] A. R. Webb. *Statistical Pattern Recognition, 2nd Edition*. John Wiley, 2002.

[26] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008.

[27] Z. Yao, X. Yang, and S. C. Zhu. Introduction to a large scale general purpose ground truth dataset: methodology, annotation tool, and benchmarks. In *Proc. of Int'l Conf. on EMMCVPR*, 2007.

[28] J. Zeng, H. Ling, L. J. Latecki, S. Fitzhugh, and G.-D. Guo. Analysis of facial images across age progression by humans. *ISRN Machine Vision*, 2012.

[29] Y. Zhang and D. Yeung. Multi-task warped gaussian process for personalized age estimation. In *IEEE CVPR*, pages 2622–2629, 2010.

[30] X. Zhu and A. Goldberg. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130, 2009.