Neurocomputing 144 (2014) 295–303

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep self-taught learning for facial beauty prediction

Junying Gan, Lichen Li, Yikui Zhai *, Yinhua Liu

School of Information Engineering, Wuyi University, Jiangmen, Guangdong 529020, China

ARTICLE INFO

Article history:
Received 6 August 2013
Received in revised form
15 April 2014
Accepted 2 May 2014
Communicated by X. Li
Available online 5 June 2014

Keywords:
Deep self-taught learning
Regression methods
Local binary pattern
Facial beauty prediction

ABSTRACT

Most modern research of facial beauty prediction focuses on geometric features by traditional machine learning methods. Geometric features may easily lose much feature information characterizing facial beauty, rely heavily on accurate manual landmark localization of facial features and impose strict restrictions on training samples. Deep architectures have been recently demonstrated to be a promising area of research in statistical machine learning. In this paper, deep self-taught learning is utilized to obtain hierarchical representations, learn the concept of facial beauty and produce human-like predictor. Deep learning is helpful to recognize a broad range of visual concept effectively characterizing facial beauty. Through deep learning, reasonable apparent features of face images are extracted without depending completely on artificial feature selection. Self-taught learning, which has the ability of automatically improving network systems to understand the characteristics of data distribution and making recognition significantly easier and cheaper, is used to relax strict restrictions of training samples. Moreover, in order to choose a more appropriate method for mapping high-level representations into beauty ratings efficiently, we compare the performance of five regression methods and prove that support vector machine (SVM) regression is better. In addition, novel applications of deep self-taught learning on local binary pattern (LBP) and Gabor filters are presented, and the improvements on facial beauty prediction are shown by deep self-taught learning combined with LBP. Finally, human-like performance is obtained with learning features in full-sized and high-resolution images.

1. Introduction

With the scientific and technological progress, people gradually begin to pay more attention to their appearances. It is human's nature to pursue beauty, and everyone longs for it. Aristotle said: "Beautiful face is a better recommendation". Whatever makes a face beautiful is a difficult question because of personal preferences causing different subjective opinions on facial beauty. Human beings have realized the existence of facial beauty since ancient times, but the ancients did not give a scientific summary on how to predict facial beauty better.

The research of facial beauty belongs to a recently springing-up leading field about human perception nature and rule [1]. At present, researchers use geometric features [2–6], apparent features [7–10] or their combination [8–10], and then predict facial beauty by machine learning methods. The former is a hotspot of research on facial beauty. Researchers extract many meaningful feature points from face images by active shape model (ASM) [11] or active appearance model (AAM) [12], compute geometric distances between feature points and ratio vectors composed of geometric distances, and then treat them as features for machine learning [2]. Eisenthal et al. [3] extracted 36-dimensional feature vectors and ratio vectors from 92 face images with beautiful and unbeautiful cases, and achieved a Pearson correlation of 0.6 with mean beauty scores labeled by humans and machine. Kagian et al. [4] utilized 90 principal components of 6972 distance vectors between 84 fiducial point locations to predict facial beauty by standard linear regression, in which face images were rated with a beauty score in 1–7 range and a Pearson correlation of 0.82 was achieved. However, simple use of geometric features to describe face images will lose much feature information characterizing beautiful faces, such as rippling muscles, structure transition of organs. Moreover, the extraction of geometric features is involved with a great deal of manual intervention, leading to no authoritative results.

Apparent features refer to the overall appearances of faces as research objects, which are not confined to simple quantity or proportion relationships, and costly manual landmarks of facial features on the analysis of beautiful characteristics. Moreover, the combination of geometric features with apparent features seems to be potentially complementary. Gray et al. [7] utilized different

* Corresponding author.
E-mail addresses: junyinggan@163.com (J. Gan), lilichen0906@163.com (L. Li), yikuizhai@163.com (Y. Zhai), yinhualiu2109@163.com (Y. Liu).

http://dx.doi.org/10.1016/j.neucom.2014.05.028
0925-2312/

sizes of filters and changed the resolution of images to extract multi-scale features of face images, and achieved Pearson correlation of 0.65. Whitehill et al. [8] employed various feature sets such as Gabor filters, Eigenface projections, edge orientation histograms (EOH), geometric features, and the combination of Eigenface projections with the others. A Pearson correlation of 0.45 was achieved by SVM regression in four-class beauty prediction, but the performance had not been improved through appending the additional features. Altwaijry et al. [9] implemented a personalized relative beauty ranking system, in which various feature sets such as histogram of oriented gradients (HOG), gist features, geometric features, color histograms, dense scale-invariant feature transform (SIFT) plus PCA, and the combination of them are employed. In this system, an average accuracy of 63% is achieved on pairwise comparisons of novel testing faces, and HOG, GIST, and Dense-SIFT+PCA features were most effective for facial beauty prediction. While acceptable results for different feature sets are produced by the methods above, simple information is just utilized to represent face images, such as edges, curves, quantity or proportion relationships, and is not involved in hierarchical, structural and high-level features corresponding to object parts, such as nose, eyebrows. What is more, the research of facial beauty by traditional machine learning methods requires a large number of training samples and a certain degree of differentiation towards beauty. That is to say, training samples should contain many very beautiful and very ugly images, but in practice most face images are neutral beauty, it is hard to meet the requirements, which undoubtedly brings a realistic difficulty to research work.

Recently, deep self-taught learning catches the attention of researchers, and empirical works already demonstrate its significance in statistical machine learning [13]. Deep self-taught learning simulates the architectural depth of brain, which processes information through multiple stages of transformation and representation, and aims at learning hierarchical and high-level features obtained by the composition of low-level features. It mainly reflects deep learning [14–16] and self-taught learning [17]. One of the efficient and popular deep learning methods is convolutional deep belief network (CDBN) [18–20], which consists of several convolutional restricted Boltzmann machines (CRBM) [21], firstly learning oriented and localized edge filters at low layer and then learns complex features corresponding to object parts at high layer. CDBN changes basic pixel into abstract concept easily accepted by human, and keeps translation-invariant representations useful to visual recognition. Lee et al. [18] combined first-layer and second-layer CDBN and achieved recognition rate of 65.4% using 30 training samples per class on Caltech101 database. However, CDBN needs a large number of labeled data to obtain good optimized starting point of network parameters, but in practice the requirement is often not satisfied. Self-taught learning automatically improves network to understand the characteristics of data distribution in the pretraining stage, and unlabeled data do not need to possess the class label or the generative distribution of labeled data. After pretraining, network just needs a small amount of labeled data to train the classifier. Raina et al. [17] utilized self-taught learning and achieved recognition rate of 46.6% using 15 training samples per class on Caltech101 database. Self-taught learning relaxes strict restrictions of the number and class of samples, which effectively solves the problem encountered by CDBN during training [18,19]. Deep self-taught learning provides more scientific theoretical basis for visual recognition, but currently there are not many researches about the concept of facial beauty and human-like predictors.

In this paper, a novel method based on deep self-taught learning is utilized for facial beauty prediction. Deep self-taught learning can effectively solve the problem of fewer restrictions on training samples, and extract structural and high-level features of samples without relying on artificial feature selection, expressing a process of automatic learning and feature extraction. In order to explore an appropriate prediction method that maps high-level representations into beauty ratings, the performances of standard linear regression [22], k-Nearest Neighbor (KNN) regression [23], ridge regression [24], multinomial logistic regression [24], and SVM regression [25] are evaluated. Moreover, the additional novel applications of deep self-taught learning on local binary pattern (LBP) and Gabor filters are presented, and the improvements on facial beauty prediction are shown by deep self-taught learning combined with LBP. Finally, much more human-like performance is obtained with learning features in full-sized and large-scale images by CDBN, which holds promise as a scalable algorithm for learning hierarchical representations from high-dimensional data.

The rest of the paper is organized as follows. In Section 2, we describe self-taught learning, and review CRBM and its stack to a deep architecture of CDBN applied to learn representations. In Section 3, we present the experiments with comparisons. Finally, we conclude the paper with discussion of future work.

2. Methodology

In this section, we first present an overview of the proposed deep self-taught learning for facial beauty prediction, and then introduce self-taught learning and convolutional RBM and DBN.

2.1. Overview of the proposed approach

In this paper, we utilize deep network to learn the representations of facial beauty. Since only a limited amount of face images are used for experiments and this may cause an overfitting problem, we employ the strategy of self-taught learning for unsupervised pretraining of deep network. Inspired by Huang et al. [19], which combined deep learning representations with hand-crafted image descriptors for face verification and achieved state-of-the-art accuracy, we present additional descriptors through learning deep network on Gabor filters and LBP, and explore whether the novel operation will help the network for facial beauty prediction or not.

Regression analysis is a statistical analysis method which identifies the quantitative relationships among two or more interdependent variables, and measures the degree that the independent variables have influence on a dependent variable. Through regression analysis, human brain is simulated and the rule of the given data is summarized and studied, thus the unknown data are forecasted. In this paper, we introduce regression analysis to predict facial beauty.

The framework of the proposed approach is shown in Fig. 1. Kyoto natural images [26] are used as unlabeled images for self-taught learning. We firstly extract LBP (or Gabor) features of Kyoto natural images, labeled face images and testing face images, and then present self-taught learning through unsupervised pretraining deep network with the help of LBP (or Gabor) features of Kyoto natural images, which limits network parameters to favorable areas for further learning and prevents deep network from falling to local minimum. After initializing deep network, we utilize LBP (or Gabor) features of labeled face images as input of deep network to train again. Training process mainly aims at helping deep network to learn and understand the structures which may characterize feature information of beautiful faces. The hierarchical features of face images can be extracted after the finish of training, and then regression analysis is used to explore the internal relationship between facial features and beauty prediction values.