Neurocomputing 177 (2016) 98–109

Contents lists available at ScienceDirect

Neurocomputing

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

Combining a causal effect criterion for evaluation of facial attractiveness models

Fangmei Chen [a], David Zhang [b,*]

[a] Department of Electronic Engineering and Graduate School at Shenzhen, Tsinghua University, Beijing, China
[b] Biometrics Research Centre and Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form
21 September 2015
Accepted 10 November 2015
Communicated by Qingshan Liu
Available online 18 November 2015

Keywords:
Causal effect criterion
Computational model
Facial attractiveness
Manipulation
Prediction

ABSTRACT

Attractiveness is an important attribute of a face that can be understood, predicted, and manipulated. Previous studies on facial attractiveness modeling focused on only prediction power. However, a model with high prediction accuracy may fail in attractiveness manipulation problems. In this paper, we add a causal effect criterion for model selection, which can be measured by imposing interventions according to the models and examining the change of attractiveness. We built a new database containing face images with diverse attractiveness and corresponding human ratings. Several manifold embedding and statistical regression methods are performed for attractiveness modeling. We compare different models under the two criteria and find that the performance of facial attractiveness manipulation depends on the causal effect of the model; feature normalization is a crucial step, without which the model will have small causal effect and fail in attractiveness manipulation; LDA manifold is better than PCA and LPP manifolds in attractiveness prediction; the linear model and the nonlinear SVR model are very similar, implying that facial attractiveness is quite directional. The selected model can be interpreted by common sense and works well for both attractiveness prediction and manipulation problems.

1. Introduction

Facial attractiveness plays an important role in one's social life. It has been demonstrated that attractive people are assumed to have more positive personalities, career success, and marital happiness [1,2]. The secrets of facial attractiveness has attracted the attention of researchers from many fields, such as psychology, evolutionary biology, aesthetic plastic surgery, and computer science. Recently, image-based facial attractiveness analysis becomes an emerging research topic in pattern recognition society, which has many potential applications, such as aesthetic plastic surgery planning, cosmetic advertising, photo editing, and entertainment.

Compared with other facial analysis tasks (e.g., face recognition, facial expression classification, age estimation, etc.), facial attractiveness study has the following challenges. First, data collection is difficult. It is partly because the face images are required to cover a wide range of attractiveness, especially sufficient beautiful faces. As far as we know, there is no public face database satisfying this requirement. And it is partly because collecting labels (e.g., human ratings) is extremely laborious and expensive. Second, due to the scarcity of beautiful faces in daily life, the images used in facial attractiveness study are often collected from the web. There are uncontrollable external variations such as illumination, occlusion, and image quality. Third, the aims of facial attractiveness study are beyond prediction, emphasizing understanding and manipulation. As shown in Fig. 1, the three modules are closely related. The understanding module can promote the researches on prediction [3–6] and manipulation [7] by supplying empirical knowledge, and the performances of prediction and manipulation modules can verify the validity of the understanding [8–10]. A valid computational model of facial attractiveness could promote all the three modules, e.g., deepening the understanding of facial attractiveness perception, estimating the attractiveness of an input face image, and guiding the manipulation on the face image to enhance its attractiveness. Potential applications of the model include cosmetic recommendation and advertisement, aesthetic plastic surgery planning, photo retouching, and entertainment.

Previous studies on facial attractiveness can be mainly divided into two categories. The first category is verification and correction of existing psychological hypotheses. Researchers manipulated the target facial characteristic while keeping other features intact to generate a treatment group and used the original face images as the control group. Then, they asked participants to judge the images in pair and analyzed the results to see if the target characteristic significantly affected facial attractiveness. The most investigated

* Corresponding author. Tel.: +852 2766 7271; fax: +852 2774 0842.
E-mail addresses: mays6633691@163.com (F. Chen), csdzhang@comp.polyu.edu.hk (D. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.11.010
0925-2312/

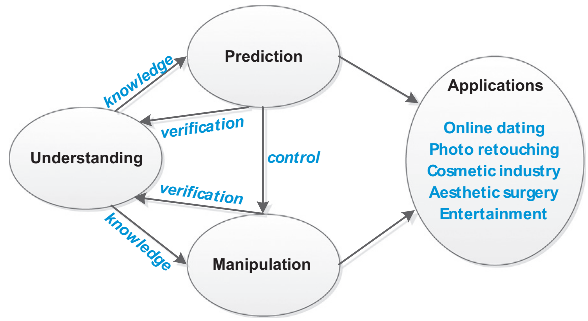

Fig. 1. Relationships between understanding, prediction, and manipulation modules in facial attractiveness study.

hypotheses are the averageness [8,11,12,10], symmetry [13,11], and sexual dimorphism [14,15,12] hypotheses and golden ratio rules [9]. These studies belong to qualitative research that only cares about whether a feature is relevant to attractiveness. They did not quantify attractiveness nor build computational models. The second category is attractiveness prediction. Different kinds of features were extracted and used to build computational models of facial attractiveness by supervised learning methods. For example, the works [3–6] used ratio features and statistical regression techniques to build models. They achieved correlations of 0.6 to 0.8 with human judgments, depending on the databases they used and the features they selected. Gray et al. [16] used multiscale local filters to extract features and adopted stochastic gradient descent method with a quadratic loss function to train the model. Nguyen et al. [17] integrated many popular facial features such as local binary patterns (LBP), Gabor filter responses, color moment, shape context, and shape parameters to build their model. In [18], the author directly used pixel values as features to learn a linear model. Gan et al. [19] applied self-taught learning to obtain effective features for facial beauty prediction. Yan [20] treated facial beauty prediction as a cost-sensitive ordinal regression problem. The aim of the above works is to achieve a high prediction accuracy, which depends only on associations. However, association is not causation. A model with high prediction accuracy may fail in interpretation and manipulation. Although [16] and [18] tried to interpret the models they learned, the learned attractive patterns they showed were not very obvious.

A model with strong causal effect is more likely to approach the true attractiveness perception rules. Knowing the causal effect, we can expect the outcomes of different manipulations on face images. However, as far as we know, there is no work considering the causal effect when building facial attractiveness models. In this paper, we add a causal effect criterion for model assessment and propose two approaches to measure it. A straightforward approach is imposing interventions according to a model and generating post-intervention face images. The post-intervention images and the original ones constitute two groups of stimuli. Then we ask participants to give judgments to the stimuli and evaluate the effect of the intervention by analyzing the judgment results. Although this approach is reliable, the cost of perceptual experiments is very high. Hence, we propose another approach to estimate the causal effect. The original and post-intervention images are represented in the feature space. Given a training set, a joint distribution between features and attractiveness can be estimated. Based on this distribution, we estimate the post-intervention attractiveness. The difference between the post-intervention attractiveness and the original attractiveness is used to measure the causal effect.

To learn computational models of facial attractiveness, a new database was built, which includes images with diverse ethnicities and large variations in attractiveness. Human ratings were collected and the mean opinion scores (MOSs) are used to measure the perceived attractiveness. A computational model contains two aspects: features and functional form. We analyze and compare different features in terms of discriminative power and manipulation convenience and select the Appearance Model [21] as the feature extractor. For convenience, we use the term AAM to refer to the model and call the parameters of the model as AAM features. AAM features combine both the shape and the texture of face images, have competitive discriminative power, and can be used to reconstruct face images by linear transformations. Hence, they are suitable for both attractiveness prediction and manipulation problems. As lack of data is a common case in facial attractiveness study, manifold embedding algorithms are applied to avoid effects of the curse of dimensionality. After that linear and non-linear models are built using linear regression and support vector regression (SVR) [22], respectively. As there are multiple potential choices in each step, e.g., if feature normalization is performed; which manifold embedding algorithm is applied; and which regression method is used, we can build different models.

Comparing different models under the prediction accuracy and causal effect criteria, we obtain some findings: the performance of attractiveness manipulation depends on the causal effect of the model. Feature normalization is a crucial step. Without this step, although the model can also obtain the same prediction accuracy, the causal effect of the model is insignificant and the model will fail in attractiveness manipulation. Manifold learning can improve the performance of attractiveness prediction, and the linear discriminant analysis (LDA) [23] manifold is better than principle component analysis (PCA) [23] and local preserving projection (LPP) [24] manifolds. The linear model and the nonlinear SVR model are nearly the same, which implies that facial attractiveness is quite directional.

Our main contributions are: (1) we build a new database for facial attractiveness study; (2) we show that the prediction accuracy criterion fails in selecting models for facial attractiveness manipulation and propose a causal effect criterion as a supplementary; and (3) the model selected by the proposed criteria performs well in both attractiveness prediction and manipulation problems.

The remainder of the paper is organized as follows. Section 2 introduces the database we built for facial attractiveness study. In Section 3, we present the causal effect criterion and propose two approaches to measure it. Section 4 describes the method of facial attractiveness modeling. Section 5 presents a model guided attractiveness manipulation method. Experimental results are shown in Section 6, and we conclude the paper in Section 7.

2. Facial attractiveness database

The face image set for facial attractiveness study is required to have sufficient variation in attractiveness. However, existing public face databases often include few extremely beautiful faces, and the faces often come from limited ethnic groups. In order to build a face database with diverse attractiveness and ethnic groups, we collected 390 celebrity face images of Miss Universe, Miss World, movie stars, and super models via the web, and collected 409 common face images from a face research website (http://faceresearch.org), Flicker, the XM2VTS database [25], and the Shanghai database [10]. All face images are confined to be female, frontal with near neutral expressions, and have adequate resolution. A pre-processing step is conducted to extract the face region and resize the cropped images to a resolution of 480×600. Human rating is a usual approach to measure facial attractiveness [3–6]. Conventionally used measurement scales are 5-point, 7-point, and 10-point scales. However, according to our experience, people find it more difficult and time-consuming to rate with more points of scale. Considering the size of our image set, we use a 3-point integer scale: -1 (unattractive), 0 (common), and 1 (attractive). An interactive tool for image rating was developed, which displayed face