AI Meets Online Dating:

User Demographics and Profile Pictures as a Predictor of Attractiveness in Facial Images

Huyen Nguyen, Duncan Wood, Tucker Leavitt

Stanford University – CS 229 Fall 2015

Overview

Although judgments of beauty are subjective, the prevalence of beauty contests, cosmetic surgery and homogeneous ratings of photos on dating websites suggest there might be a common denominator for attractiveness perception. This project is an attempt to quantify that denominator.

In this project, we explore the notion of facial attractiveness using machine learning. Our model attempts to predict attractiveness scores for facial images. The input for our model is a frontal image of a female subject. The output will be a discrete value between 1 and 5 which represents the attractiveness of the subject in the image.

Through this project, we want to achieve two goals:
1. We want to see how our model generalizes with racially diverse datasets. For example, we want to train our model of a data which consists solely of Asian females and test our hypothesis on datasets of Caucasian or African females.
2. We want to train our hypotheses on a racially diverse dataset to see if it will increase or decrease the correlation.

Background

We use three different datasets: the SCUT-FBP dataset, Chicago Face dataset, and images scraped from the HotOrNot dating website.

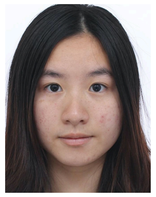
The SCUT-FBP dataset contains 500 frontal images of 500 different Asian females with neutral expressions, simple backgrounds, and minimal occlusion. These factors are conducive to facial beauty perception in both geometry and appearance. Each image is rated by 75 users, both males and females, on a discrete scale from 1 to 5. We computed the average rating score for each image and used these as the labels.

The Chicago Face Dataset includes high-resolution frontal photographs of 597male and female targets of varying ethnicity. Each target is represented with a neutral expression photo that has been normed by an independent rater sample. Each photo is rated by 20-30 raters with the average rating being a continuous value between 1 and 5.

The HotOrNot contains scraped from HotOrNot website. Each image has a continuous rating from 0 to 10, which is the ratio between the number of people who like her and the total number of people who rate her profile as like/dislike. For example, if 8 out of 10 people like a person, that person would have a rating of 8. We scaled the ratings of HotOrNot dataset to [1,5] range so that they can be the same as the ratings of the SCUT-FBP dataset and Chicago Face dataset. This dataset is used for testing purpose only.

Due to time constraints, we only process frontal images of female subjects.

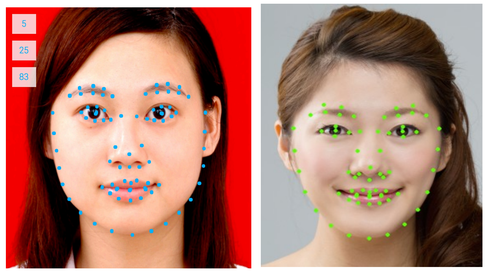
An subject from Chicago Face Dataset with the attractiveness rating of 4.36



An subject from SCUT-FBP Dataset with the attractiveness rating of 2.49

Methods

1. Feature Extraction
The model is trained two kinds of features: landmark features and spatial features.
Landmark features are (x,y) coordinates of important points on the face. We used Facemark API to extract 68 different important points on a face.

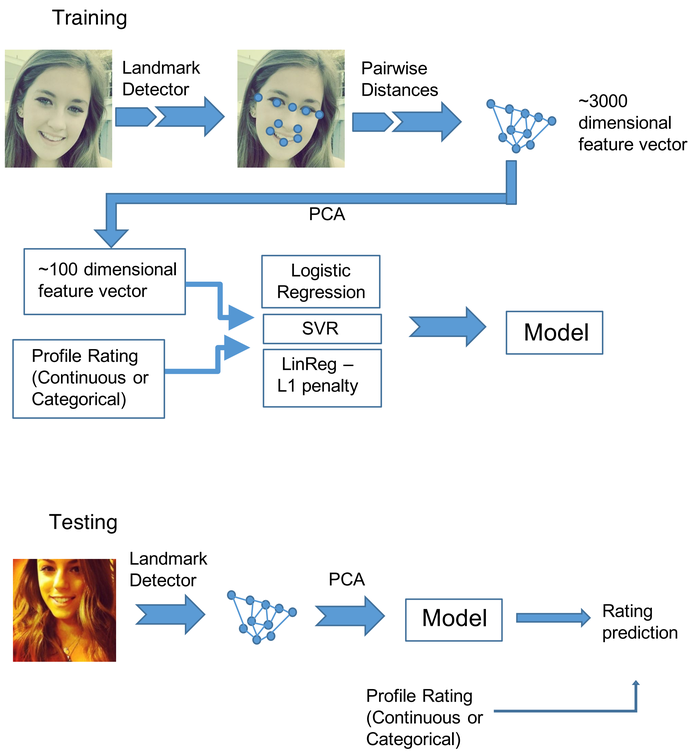


Spatial features are the distances between any two landmark points on the face. They are normalized as compared to the width of the face. The number of spatial features are 68 choose 2 = 2278.

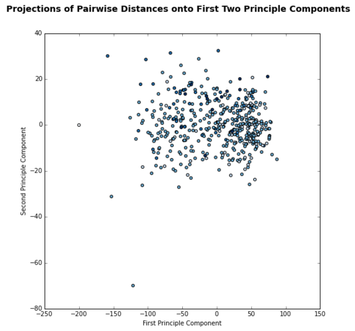
2. Dimensionality Reduction
Due to the large amount of features, we use Principal Component Analysis (PCA) to reduce the dimension of our data to 100 features.

3. Algorithms
We use Python's scikit-learn package. For now, we train our models using two different algorithms: Multiclass Logistic Regression with L1 regularization, and Support Vector Regression.

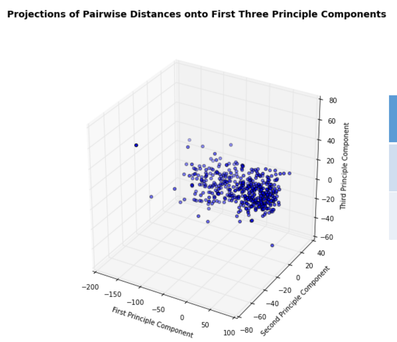
Model Overview



Results

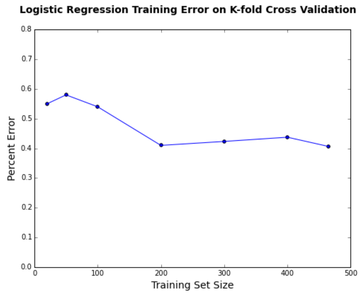


Darker colored points correspond to higher rated profiles



So far, we have been able to achieve the following classification accuracies:

Model	Classification Accuracy
Logistic Regression – L1 Penalty	60% (percent correct)
Support Vector Regression	0.71 (coefficient of determination)



Training set error was determined using k-fold cross-validation with k = 0.1*(training set size)

Conclusion

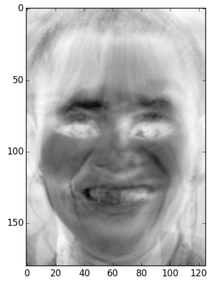
We found that based solely on the set of distances between key points on a photo, we can predict with as much as 70% the score on a scale from 1-5 of the attractiveness of a woman's face as rated by males. This result is extremely significant considering the subjectivity involved in determining attractiveness, in addition to the many features not captured by this model including complexion and hair and eye color.

This means that for women choosing which photos to post online, the way the face is depicted spatially can be a principally important consideration.

Next steps

A more detailed analysis on the Principal Components may prove to be interesting. For instance, from a given principal component vector, what would it look like to reconstruct a face from different points in the lower-dimensional feature space? Do they correspond to positioning of the head, or more structural properties?

We have found more photos with ratings, which we plan to add to our dataset and analyze with the same methods. We hope this will further increase accuracy, and since the data is not restricted to Asian women, help with analysis of more general properties of attractive faces. This will tell us whether the properties that make women attractive are the same across ethnicities.



Another avenue to pursue:
The "Eigenfaces" method, from running PCA on the photos directly