EigenHot or EigenNot

A personal preference learner for female attraction to males based on facial symmetry, masculinity index, and descriptive booleans, utilizing Supervised K-Means.

Noah Youngs

1. Motivation

In the age of social networks, more and more couples meet online, many through dating websites. Several of these sites advertise "algorithms" that match users with potential matches based on a set of self-reported personality attributes. Physical attraction is another important component of a successful relationship, however, and it stands to reason that if a person's preferences are somewhat consistent, then it should be possible to teach a classification algorithm to match those preferences based on past photograph ratings.

2. Related Work

In 2005, a similar idea was proposed by Eisenthal, Dror, and Ruppin, in their paper: "Learning Facial Attractiveness". The paper attacked the problem from two different directions: A PCA decomposition along the lines of the Eigenfaces facial recognition algorithm developed by Sirovich and Kirby, and also a linear kernel SVM applied to a set of 37 features that were manually mapped (such as distance between eyes, average facial tone, etc.). Through a hybrid of the two algorithms, a 65% correlation was achieved with the attractiveness score of each image as determined by a panel of human graders.

In order for the PCA analysis to generate meaningful results, the images were first aligned and scaled to insure that key landmarks on each face lined up. PCA was also applied to the manual feature space along with SVM, with intermediate eigenvalues showing the largest correlation with attractiveness. Several different kernels were tried with the SVM application, but none showed any significant promise above the linear.

Eisenthal et al. concluded that the largest detriment to the algorithm's success lay in the small dataset, showing by way of evidence a plot of the increasing correlation between the hybrid algorithm and human graders as the sample size increased.

3. Data

Face photos were obtained as a subset of the Put Face Database from CIE Biometrics [5]. Of the 100 subjects, 84 were male, and 16 female. Since the preferences for attraction vary across the sexes, and the male subjects were far more numerous, the female photos were discarded. The remaining 84 photos were of men approximately 18-40, all with neutral expressions, looking straight at the camera.

Because the data set was originally comprised of action shots taken of a head turning, there are slight variations in the angle of the head with respect to the camera, as well as distance to the camera. The photos are in color, and of dimension 1536x2048 pixels. Additionally, the pixel coordinates of 20 major facial landmarks were recorded by the database creators, as pictured in Figure 1.

4. Methodology

Utilizing several psychological studies on the most important male facial features and relations in a women's determination of his attractiveness, a feature set of 15 measurements was extracted using the landmarks provided by the dataset. These included the masculinity index [6], vertical and horizontal symmetry measures [6], the area of the eyes and the length of the chin [2], a few other length and width rations, as well as a skin tone measure obtained by averaging patches on the cheeks and forehead converted to grayscale. To these features were added 5 boolean variables for eye color, hair color, hair length, and the presence of a beard and/or a moustache.

In addition to the 84x20 feature set, another data set was used consisting of the pixel-space representations of every image. These were converted to grayscale, and the original images were cropped down to a size of 1000x1100 pixels. The images were then rescaled to 6% of their original size for the sake of computational costs, and then the covariance of the resulting 3960 element vectors was computed. The largest 20 "Eigenfaces" (see figure 2) were used a basis for the space, and then several classification algorithms were used on both the feature matrix and the post-PCA pixel matrix.

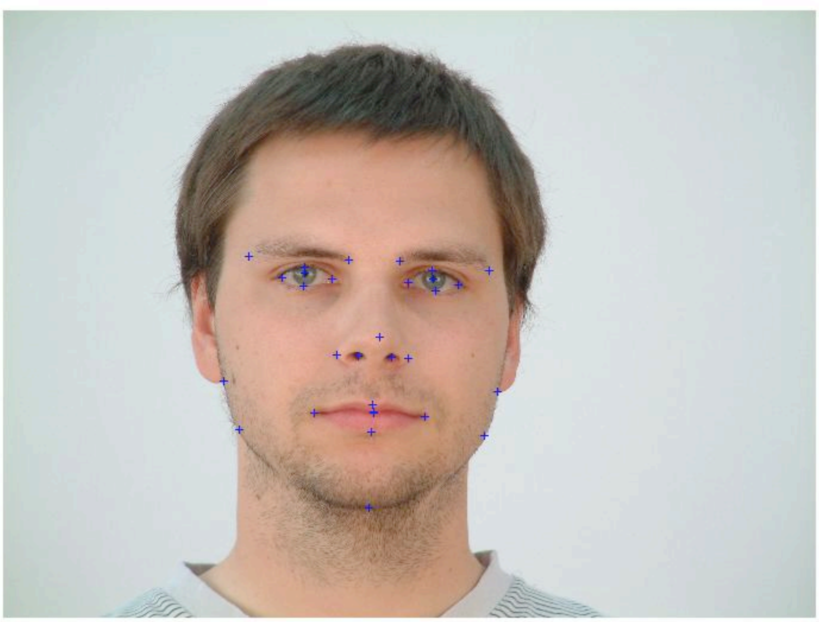

Figure 1: Facial Landmarks

Figure 2: Top 20 Eigenfaces

The classification for each image was a binary label corresponding to whether or not the rater found the face in a particular photo "attractive" or not. Three different women, aged 18-22 separately rated each subject in the database, going through the photos twice in different orders, and a third time if necessary to break any ties.

The first algorithms run on the data sets were SVMs with linear and Gaussian Radial Basis Function Kernels (SVM code courtesy of [3]). To ensure each feature of the data was on a similar scale, the features were normalized to unit variance. Subsequently K-Means was run on the data for cluster sizes running up to 42, with the label of each cluster determined by the most prevalent label of the objects within that cluster.

The successes of both algorithms was measured by Leave-One-Out Cross-Validation error, since the small sample size was prohibitive of further splitting the data into training and test sets. In an attempt to overcome the sensitivity of K-Means to local minima, during each iteration of the Cross-Validation K-Means was run 10 times (due to computational constraint) with different random initial centroids, and the cluster set with the smallest classification error was chosen.

In addition to these basic algorithms, the modified Supervised K-Means Algorithm [1] was run on the data sets in an attempt to reduce the classification error. Supervised K-Means utilizes the weighted-Euclidean norm:

$$\delta_{x,y} = \sqrt{\sum_i w_i (x_i - y_i)^2}$$

Where the weight vector W is determined in order to segregate the resulting K-Means clusters by class-type as much as possible. This is accomplished by defining an objective function as the number of objects with a different label than the predominant label of the cluster they are in, and then choosing W to minimize this function. As proposed by Al-Harbi and Smith, this minimization was accomplished by Simulated Annealing. A cooling parameter of 1 was utilized, with a multiplicative factor of .95 applied every 100 iterations of the annealing minimization.

5. Results

The Cross-Validation errors for the support vector machines on both data sets are summarized in table 1. The type of machine had no effect on

Data/Kernel	**Cross-Validation Error**		
	1	**2**	**3**
Feature/RBF	**23.81%**	**38.1%**	**23.81%**
Feature/Linear	**23.81%**	**38.1%**	**23.81%**
Pixel/RBF	**23.81%**	**38.1%**	**23.81%**
Pixel/Linear	**23.81%**	**38.1%**	**23.81%**

Table 1: SVM Results

the cross-validation error. This is because in all cases, the SVM's classified the test subject as "not attractive". The first and third reviewer classified 20 of the 84 subjects as attractive in the initial ranking, and the second reviewer classified 32 as attractive, which accounts exactly for the Cross-Validation errors observed if the SVM were to declare universal "unattractiveness".

The K-Means clustering with cluster labeling was able to perform slightly better for several cluster sizes, as shown below in figure 3.

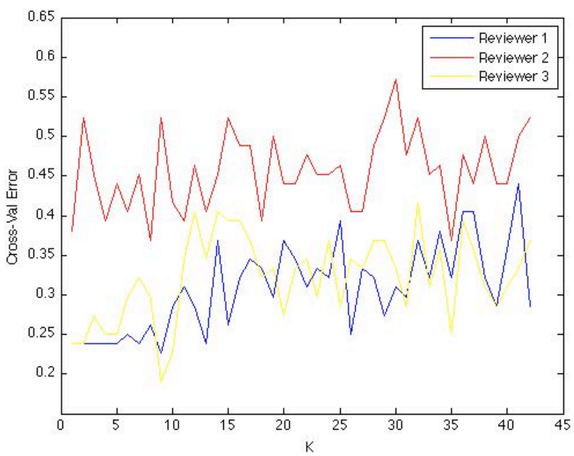

Figure 3: K-Means Cross-Val Error

For reviewers 2 and 3, the K-means clustering algorithm beats SVM, but only for a few cluster sizes and only slightly. When the Supervised K-Means algorithm is implemented for a cluster size of ten, the classification error was reduced for all three models by a few percentage points. Ideally, new weights would be computed for each cluster size, but due to the computational complexity of the weight calculation, the same maximal weights from K=10 were applied to all cluster sizes in the Cross-Validation testing. Unfortunately the resulting errors only beat the unsupervised K-Means for a few cluster sizes for some of the reviewers. The results of Supervised K-means are shown in figure 4, and Supervised and unsupervised K-Means for Reviewer 2 are compared in figure 5.

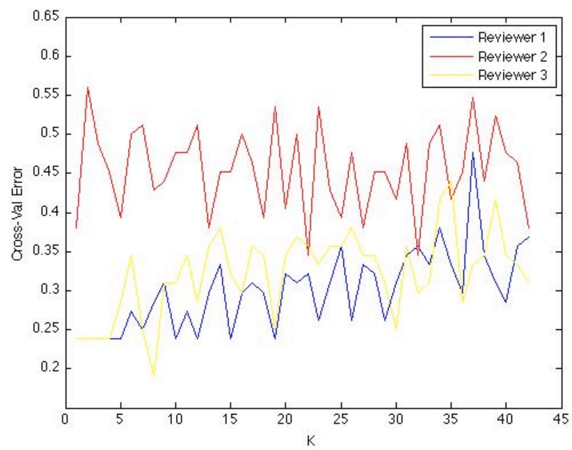

Figure 4: Supervised K-Means

Both of these graphs show the results on the feature data. The results of these algorithms applied to the PCA-reduced pixel-space are similar in quality, and have been omitted due to length constraints.

6. Conclusion

The lack of striking results is most likely due to three main causes: quality of the data set, psychological factors, and computational constraints.

The data set was limited in size, and also featured many similar-looking subjects. In addition, the landmarks provided by the dataset were not always exact, and were also not quite the same as the landmarks needed to create the

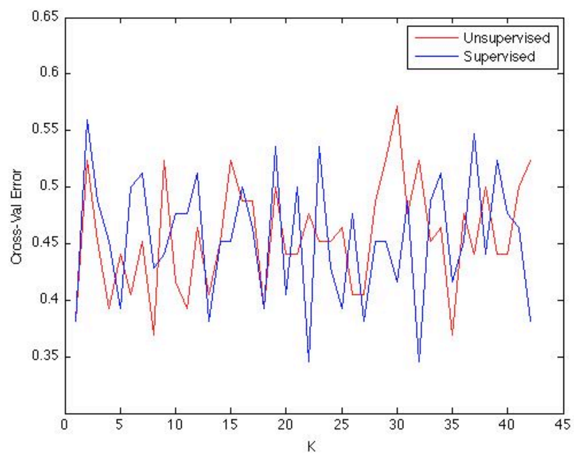

Figure 5: Supervised vs. Unsupervised, Reviewer 2

masculinity ratio. Thus the ratio used was an approximation to the masculinity ratio, and prone to small variation in the accuracy of its measurement. Also, the rotation and scaling of the faces were prone to minor variation, which could have had an impact on the eigenfaces decomposition. Finally, the reviewers all agreed that the majority of the subjects fell into the unattractive category, leaving a lack of examples of attractiveness for the algorithms to learn from.

In terms of psychological factors, the feature set used relied on psychological conjecture, since the true nature of attraction is not fully understood, and therefore a different set of features, or additional features, might yield better results. The reviewers also seemed to express "sympathy" while rating the photos, implying that some of the "attractive" classifications were out of pity, rather than indicating true belief, which would further muddy learning attempts. Finally, many studies agree that women are more prone to circumstance and personality cues in their attraction to men, rather than static facial features, and thus if the genders of the problem were switched, better results might ensue, but this would require a new data set as well.

Computational constraints for the problem included the restriction of the size of the images when calculating the eigenfaces, as well as on the number of times K-Means was run for each cluster size, creating sensitivity to local minima. The time frame of the Simulated Annealing for Supervised K-Means was also restricted, and the small sample size made for a step-function like objective function, further hampering the search for good weights.

With a gender switch, larger and more representative data set, and more computational power, the prospect of personal preference for attraction learning is still a reality.

7. References

[1] Al-Harbi S, Rayward-Smith VJ. "Adapting k-means for supervised clustering". Applied Intelligence 24.3 (2006): 219-226.

[2] Cunningham M, Barbee A, Pike C. "What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness". J Pers Soc Psychol 59.1 (1990): 61-72

[3] Eisenthal Y, Dror G, Ruppin E. "Learning Facial Attractiveness". Unpublished. 2004 draft available at: http://www.cs.cmu.edu/~cga/behavior/faces1.pdf

[4] LS-SVM lab: http://www.esat.kuleuven.be/sista/lssvmlab/

[5] Put Face Database, courtesy of Adam Schmidt and CIE Biometrics: https://webmail1.cie.put.poznan.pl/biometrics/index.php?option=com_content&view=frontpage&Itemid=1&lang=en

[6] Scheib J, Gangestad S, Thornhill R. "Facial attractiveness, symmetry and cues of good genes." Proc Biol Sci 266(1431) (1999): 1913-1917.