

Gender and Facial Expression Classification (Intermediate project report)

Jifu Zhao

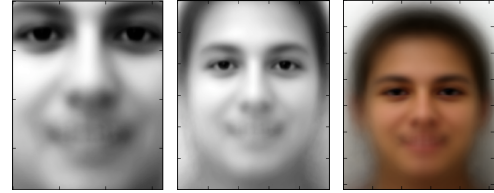
Graduate Student

Department of Nuclear, Plasma, and Radiological Engineering

jzhao59@illinois.edu

1 Project Review

In this paper, the data is from FEI Face Database (Thomaz and Giraldi, 2010). There are 200 individuals including 100 males and 100 females. Both have two pictures with smile and non-smile expressions. So, this problem is basically a multi-class classification problem.



(a) frontal face (b) full face (c) colored face

Figure 1: Mean faces

2 Progress made

After collecting the data, first they were correctly labelled manually into 4 classes-man smile, man non-smile, woman smile, and woman non-smile.

Then, using Python code, this original data set is randomly shuffled. And we select 75% (300 faces) as the training set, and the rest 25% (100 faces) as the test set.

In this question, the feature for three different pictures (frontal faces in gray scale, full faces in gray scale, colored faces) are as follows:

$$\begin{aligned}193 \times 162 &= 31266 \\300 \times 250 &= 75000 \\360 \times 260 \times 3 &= 280800\end{aligned}$$

Based on the feature of the picture, a significant part of the original feature space is not useful for us. And if we do classification in the original feature space, we need large memory and powerful CPU. We don't want to waste resource on this, so our first step is to decrease the dimensionality. Here, we use Principal Component Analysis(PCA) method.

In PCA method, the first step is to center the data. After center the data, the mean faces are shown in Figure 1.



Figure 2: Original frontal faces



Figure 3: Centered frontal faces

After center the data, the original face will change a little bit. The original faces and centered faces are shown in Figure 2 and Figure 3 (Since the following analysis method are the same for three group figures, I will only show the result for **frontal faces**).

Through PCA analysis, we can get a series of eigenvectors. In fact, there eigenvectors are the eigen-faces. They are shown in Figure 4.



Figure 4: Eigen frontal faces

The corresponding explained variance ratio, which measure the relative significance of the eigen faces, are shown in Figure 5

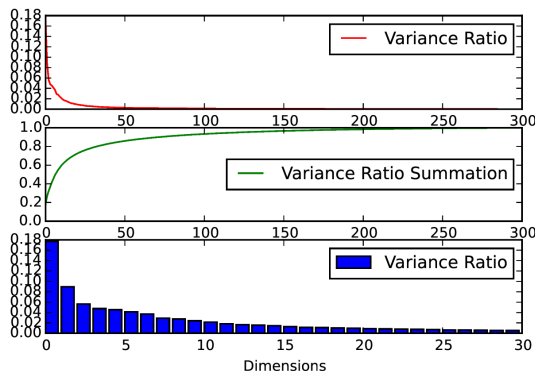


Figure 5: Explained variance ratio

As, you can see from Figure 5, although there are originally 31266 features, after PCA, 100 dimensions will include about 90% changes. This means that our following analysis can based on the reduced feature space.

After PCA, we in fact got a series of eigenvectors, which stands for the eigen-faces. These

eigen-faces can be used to reconstruct the original faces. This means that if we can maintain most of the information only use first several dimensions. For example, if we use only the first 2 dimensions and project the original faces onto these two eigen-vectors, we can get the result shown in Figure 6.

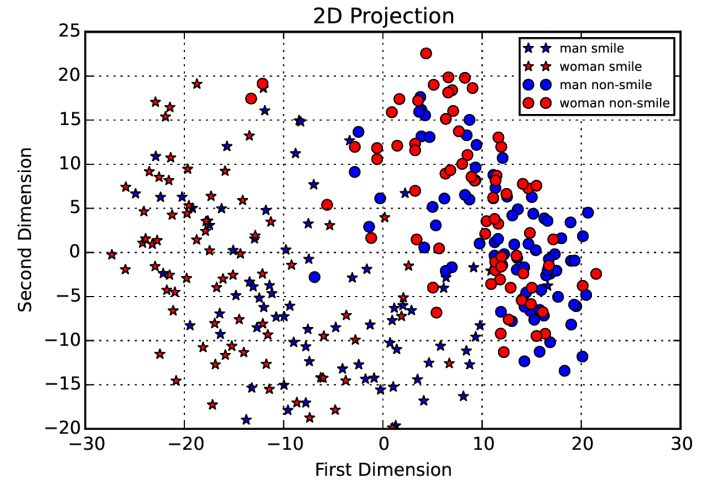


Figure 6: 2 dimension projection

As you can see in Figure 6, although smile and non-smile faces are not separated very well, man and woman faces have a relative very good separation. This is because that we only use the first 2 dimensions. The next step is to use more dimensions and try more algorithms on this problem.

3 Plans

Until now, I have successfully reduce the dimensionality from more than 30000 in to about 100 dimensions. The next step is to try some classification algorithms, such as support vector machines (SVM), Perception, Quadratic Discriminant Analysis and so on. Our goal is to quantitatively analyze this problem.

4 Difficulties

Currently, the only problem is the data set problem. Now, we only have 400 faces. But we have more than 30000 features. This will influence the accuracy of our analysis. But we are still looking for more data.