QEA Project 1: BOW Physiognomy

Sam Coleman, Caitlin Kantor

February 2020

1 Summary

This report will investigate the ethics and efficacy of using facial recognition technology to categorize students. Using Eigenvector decomposition and principal component analysis, we test if it is possible to accurately determine if a student attends Babson, Olin, or Wellesley (BOW) based on their LinkedIn profile picture. The introduction provides insight into the ethics surrounding our question and introduces the algorithms we used. The methods section goes into more detail about the steps we took and the mathematical concepts that are embedded. The detailed findings section explains the outcome of our program and compares the efficacy of the various approaches we used, along with the consequences of our program. Overall, our first method is not very accurate (max accuracy is 44%) because of the various positions students in the pictures are in such as showing torso, full body, just face, etc (see figure 5 and figure 6 for examples).

2 Introduction

2.1 FeaRDeClass Background and Context

We have decided to investigate the use of facial recognition technologies in society. There is a broad context in which this technology is applied, both on the micro and macro scales. Facial recognition has become an integral part of people's daily lives—from unlocking your phone by using your face to Google Photos grouping images by person to China's social credit system. The former seems harmless, as your data remains local on your phone and is only used to identify whether the face is or is not yours. However, there is always a chance that the data will be hacked and taken into the wrong hands. The latter seems more problematic as China is tracking their citizen's every action and giving them a corresponding score. Actions China deems positive adds points to a user's social credit, while actions China decides are negative decrease points. These scores are then used to reward ideal citizens and punish those with low scores. An integral part of this system is using facial recognition to track all of its citizens.[1] This technology can also be applied to situations beyond faces, such as to identify objects. Additionally, these same algorithms can be used to profile groups of people. This is currently happening in China, where a specific facial recognition system is exclusively looking for Uighurs based on their appearance and stores records of their movements for review. The number of provinces requesting this technology is increasing as more police departments want to utilize the profiling of this group of citizens.[2]

2.2 Technical Details of Algorithm

Facial recognition algorithms use linear algebra to detect faces that are most correlated to the original face they were trained with. To start this process the algorithm is first given a large matrix that contains a set of training data containing pictures of faces. Each column as a whole is a representation of an image with each row containing one pixel. Each pixel that falls in that row across the columns represents the same pixel location in all the images. The rows are the pixels of each image in a column vector and the columns are the number of photos. The algorithm then analyzes all of the pixels that make up each image to determine where

[1] https://www.wired.com/story/china-social-credit-score-system/

[2] https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

the data is most variant. This is done by using a covariance matrix, which is an unnormalized version of the correlation matrix that does not divide each value by the standard deviation. Thus, we preserve the intensity of each pixel in the matrix. Using the covariance matrix, we can determine each way the data is varied. This results in different "eigenfaces." The eigenfaces are eigenvectors that are based on the eigenvalues. Each eigenface represents a variation in a different dimension that is normal to all other eigenfaces. The square root of the eigenvalue associated with that eigenface tells us the magnitude of variation in that dimension. The eigenface with the most amount of variation is called the Principal Eigenvector. Any scaled combination of these eigenfaces results in a different face. The process of breaking faces up into different eigenfaces is called Eigenvalue Decomposition. Since this is true, the algorithm can take any image of a face, deconstruct it into different eigenfaces, and then compare each eigenface in its respective dimension. In many situations, you can end up with more eigenfaces than are actually useful. In this case, we use Principal Component analysis to choose eigenfaces we actually use. Typically, we want to use the eigenfaces with the most variance, or the largest eigenvalues, thus the name "Principal Components."

2.3 General Ethical Implications of BOW Physiognomy

We acknowledge the ethical issues that are raised when taking our application out of this strictly educational context. While students among the BOW schools regularly will identify which students are "one of us vs one of them," putting it into an algorithm raises much larger issues. There is already a culture of Babson Vs. Olin Vs. Wellesley across the campuses, and this would just further the divide. There is a difference between thinking you are different from students at these institutions and seeing a differnece through an algorithm. The added level of identification through an algorithm insinuties that there is an innate difference between students, which would likely amplify stereotypes such as "Babson students are jocks" and "Olin students look like nerds." Sorting people into different subjective categories based on looks is not an issue to be taken lightly and is problematic regardless of the situation.

2.4 Question and Sub-question

The big question this paper will investigate is the feasibility of identifying which BOW institution a student attends based on their appearance. To answer this, we will be exploring how we can use Eigenvalue Decomposition and Principal Component Analysis to accurately determine which BOW institution a student attends based on their LinkedIn profile photo. We will analyze how accurately our algorithm correctly places students into their respective schools by using a testing set of LinkedIn pictures.

3 Method

3.1 Method 1

In our first method of answering the presented question, we used the full LinkedIn profile pictures and utilized Eigenvalue Decomposition and Principal Component Analysis. Our training data consists of photos of Babson, Olin, and Wellesley students. We took the transpose and then mean centered the photos within each school. We converted these sets into column vectors, where each row is the full representation of an image and the number of columns represents the number of samples (images). Next, we combined these matrices of Babson, Olin, and Wellesley into one large matrix. The test data is a matrix of new Babson, Olin, and Wellesley photos in the same format, along with an "answer matrix" with a value representing the corresponding photo's school. This matrix is also mean centered. The following steps are critical points of our algorithm:

1. Find the covariance matrix of the column vector
   (a) $A = \frac{1}{\sqrt{N-1}} \times [MeanCenteredPhotoMatrix]$, where A is the scaled and mean centered matrix, N is the number of images (rows), and MeanCenteredPhotoMatrix is your training data image matrix as described in the above paragraph with the dimensions $[N, x \times y]$ where $x \times y$ is the dimensions of each photo.

(b) $R = A' \times A$, where R is the covariance matrix

2. Find the first n number of eigenvectors in our training based on n number of principal components. The eigenvectors represent the directions in which our data has the most variance, within n dimensions. This compresses our data by only focusing on the dimensions with the most variance. These are the eigenvectors that have the largest corresponding eigenvalues.

3. Project the data into "bodyspace." This is done for both the training and test data.

   (a) $body\_space = [MeanCenteredColumnVector] \times EigenVectors$, where MeanCenteredColumnVector is done for both testing and training data and Eigenvectors are determined in step 2

4. Find the nearest neighbor for each row in the testing data. This finds the corresponding photo in our training data that has the least distance between the points.

5. Determine the accuracy of our data by comparing each result to the answer matrix and finding the average.

4 Detailed Findings

The results of our first method are inaccurate. While our model is more accurate than if you were to randomly assign a student to one of the three BOW schools, it is not significant enough to produce reliable findings. As we expected, the number of principal components affected the accuracy of our results. While we have some outliers, especially one principal component giving the highest accuracy of 44%, while using the maximum number of principal components (225) gives an accuracy of 40%. Disregarding outliers, the general shape of our accuracy plot follows the expected curve (Figure 1).

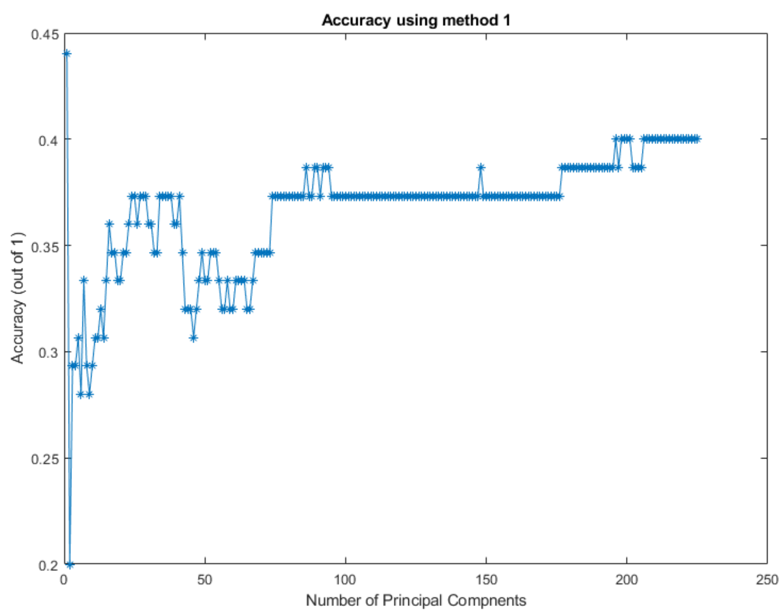

Figure 1: Method 1 Accuracy Plot

Figure 2 displays the number of students sorted into each institution by our algorithm compared to the institution they actually attend. Label 1 represents Babson, label 2 represents Olin, and label 3 represents Wellesley. The diagonals starting from the top left show that our algorithm correctly classified 30 students (out of our testing data of 75). Our algorithm is most accurate categorizing Babson and Olin students, which is interesting given that both institutions accept students of all gender identities. Wellesley, on the other hand, have the policy "Wellesley will consider for admission any applicant who lives as a woman and consistently identifies as a woman. Therefore, candidates assigned male at birth who identify as women are eligible to apply for admission. The College also accepts applications from those who were assigned female at birth, identify as non-binary, and who feel they belong in our community of women. Those assigned female at birth who identify as men are not eligible for consideration for admission".[3] Babson and Olin both have an approximate gender breakdown of 48% female and 52% male.[4,5] Based on this, we originally thought that our algorithm would be most accurate classifying Wellesley students, however, this is not the case. Furthermore, although we have vast inaccuracies across all three institutions, the greatest inaccuracy comes from incorrectly identifying Wellesley students as Babson and Oliners. In fact, more Wellesley students were classified as Babson students than were correctly placed into Wellesley. This could be because of the poses students at each school chose for their profile pictures. Furthermore, the highest misclassifications of Babson students was into Wellesley. Olin students get most often misclassified as Babson students. These relationships are extremely interesting and could be further investigated in the future.

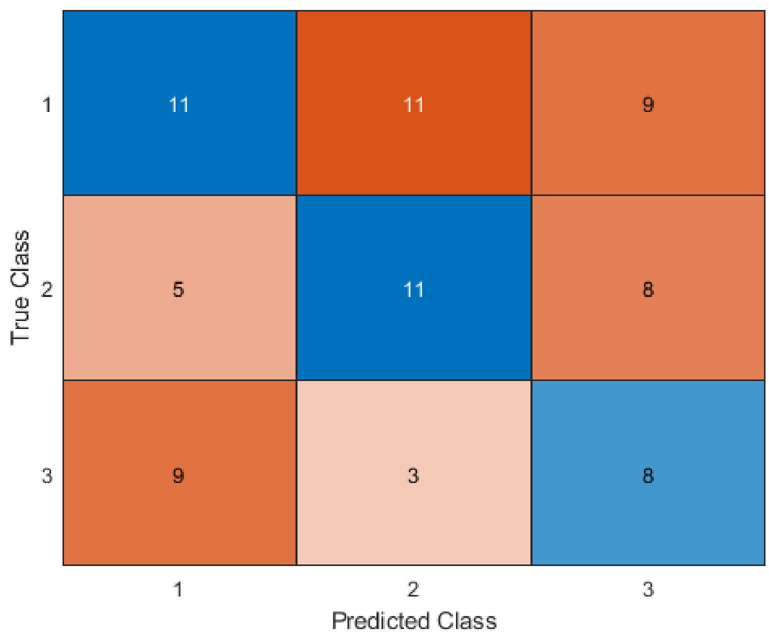

Figure 2: Confusion Matrix

A significant factor in the inaccuracies of the first method lies in the differences of the photos used in the training and testing data. Each person has complete control over the photo they choose for their profile picture, creating a large amount of noise in the photos. Not only does each photo have a different amount of lighting, but also various angles, portions of body displayed, etc (Figure 4). This resulted in noisy eigen bodies produced, and created body position as a main part of our algorithm (Figure 3). The hazy appearance created by the overlaying of different images in the eigen bodies are indicative of the various poses people chose.

[3]https://www.wellesley.edu/news/gender-policy
[4]https://www.usnews.com/best-colleges/babson-college-2121: :text=
[5]http://www.olin.edu/about/at-a-glance/

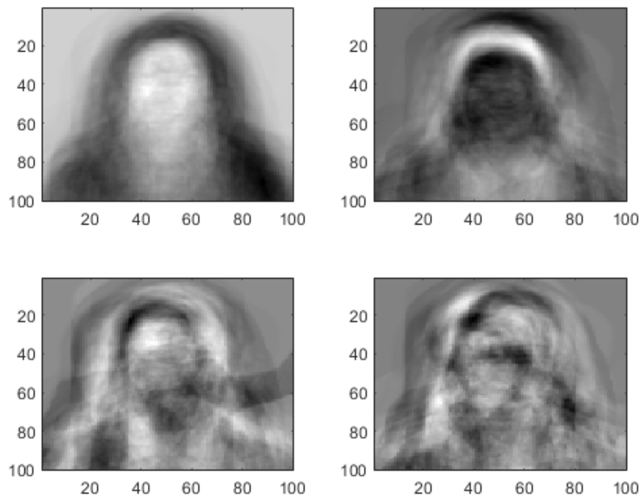

Figure 3: Sample Eigenbodies Produced

Figure 4: Sample LinkedIn profile pictures in training data

By not cropping the faces out of the images for method one, we edited our training and testing data the least. This resulted in our algorithm taking into account the various poses, as we wanted to allow the student's choices in their profile picture to influence what school our algorithm sorts them into. After analyzing our data by displaying what image in our training set each testing photo was assigned to, we were able to clearly see this correlation. As (figure 5) shows, the test image (right) was placed with the training image (left) based on the amount of torso shown. Both students show about the same amount of their body, are overall centered in their frame, and body proportions are similar. Another example is shown in Figure 6. Both students are centered in their frame, both show shoulders and head, have a straight shot camera angle, and heads take up about the same amount of the frame. This shows that our algorithm matched students based on body pose and position.

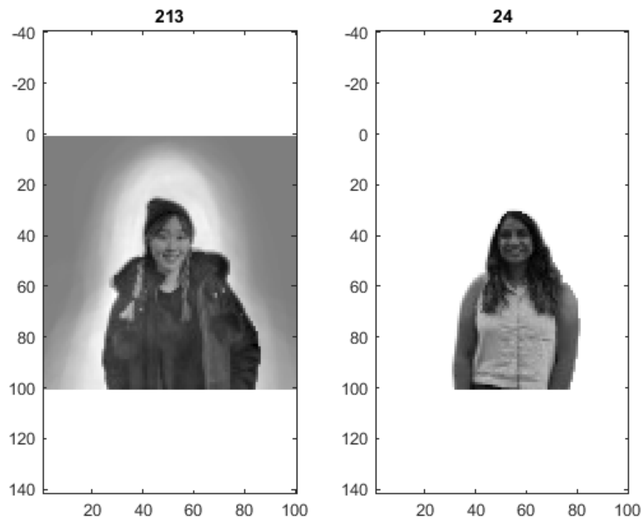

Figure 5: Framing example based on torso shown

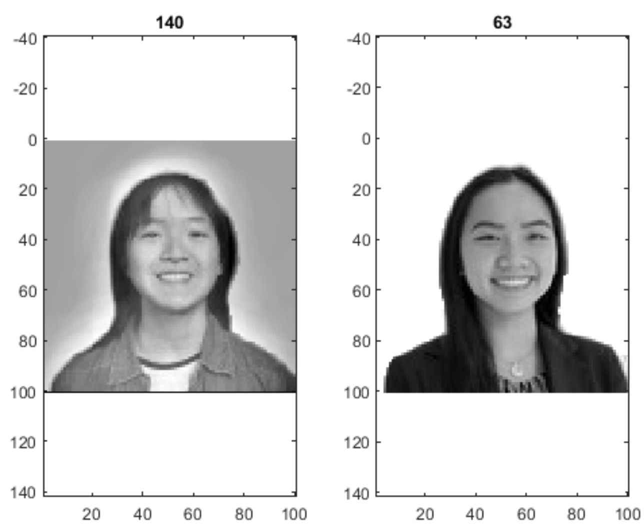

Figure 6: Framing example shoulders and up

A major downside of this technology is the blatant classification of people based purely on their looks. There are multiple groups of people this can hurt. The first are the current students of Babson, Olin, and Wellesley. This technology reinforces stereotypes in society to a great extent. With all the BOW wide events taking place throughout the semester, in addition to the existence of the Weismann Foundry for the collaboration of students across all three campuses, this technology could result in a large step back if used incorrectly. It can reinforce the thought that Babson, Olin, and Wellesley students are extremely different from each other. Furthermore, it could also hurt incoming students. As Olin is an engineering school, someone who gets classified into Olin "looks like an engineer." If a student is striving to become an engineer and were to use this dataset, they may feel a huge hurdle if they get classified into Babson. This would reinforce the "you don't look like an engineer" stereotype.

5 Recommendations

Overall, our model is rather inaccurate. We believe this mainly stems from the fact that there is a lot of noise created by different angles as well as the type of photo people chose for their profile picture (e.g. headshot, full body, etc) that made up our initial training data. One way we could improve our results is by cropping the images so that they just focus on the faces of each student. However, this could take away from the fact that students from each school may "style" themselves in similar ways. Furthermore, we realize that by including Wellesley, a historically women's college, in our training data, we create a group where nearly 100% of the photos are of people that have traditionally feminine features. This when compared to more of a 50/50 split at Babson and Olin causes our model to sort Wellesley students into other schools. One way we could try to improve this is by excluding Wellesley and just focusing on Babson and Olin which are closer in gender demographics. This would theoretically lead to more accurate results. We could also focus on the women at each respective school but that would drastically decrease the demographics of where our model applies.