Linear and Generalized Linear Models for Analyzing Face Recognition Performance

J. Ross Beveridge
Colorado State University
Credit Where Credit is Due …

- Bruce Draper …… CSU Computer Science
- Geof Givens ……… CSU Statistics
- Jonathon Phillips …. NIST
- Graduate Students
  - Wendy Yambor, Kai She, David Bolme, Kyungim Baek, Marcio Teixeira, David Bolme, Ben Randall, Trent Williams, Jilmil Saraf, Ward Fisher
What Factors (Covariates) ?

- Race
- Gender
- Age
- Eyes
- Glasses
- Bangs
- Mouth
- Facial Hair
- Smiling?
Subject Image Data
Yes, Yes, FER(R)ET Again …

http://www.rolmop.org/ferrets/

Ross Beveridge, Biometric Quality Workshop, March 9, 2006
Subject Image Data

- 1,072 Human Subjects from the FERET Data
- 2,144 FERET Images
- Exactly 2 images per subject, taken on same day
Collecting the Covariates
# Our Subject Covariates

<table>
<thead>
<tr>
<th>FERET Subject/Image Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Per Subject</strong></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>Skin</td>
</tr>
<tr>
<td><strong>Fixed Per Image</strong></td>
</tr>
<tr>
<td>Bangs</td>
</tr>
<tr>
<td>Expression</td>
</tr>
<tr>
<td>Eyes</td>
</tr>
<tr>
<td>Facial Hair</td>
</tr>
<tr>
<td>Makeup</td>
</tr>
<tr>
<td>Mouth</td>
</tr>
<tr>
<td>Glasses</td>
</tr>
</tbody>
</table>
Standard Algorithms to Test

Three Algorithms

- PCA
- IIDC
- EBGM

http://www.cs.colostate.edu/evalfacerec/index.html

Ross Beveridge, Biometric Quality Workshop, March 9, 2006
NIST FERET Image Preprocessing

- Integer to float conversion
  - 256 gray levels to single-floats
- Geometric Normalization
  - Human chosen eye centers.
- Masking
  - Elliptical mask around face.
- Histogram Equalization
  - Equalize unmasked pixels
- Pixel normalization
  - Shift and scale pixel values so mean pixel value is zero and standard deviation over all pixels is one.

Refinement of NIST preprocessing used in FERET.
Training

- **Best, but infeasible, solution**
  - Disjoint images, same set of human subjects.
  - But, subject replicate images limited in FERET.

- **Next best choice**
  - Train on exactly those images used in the study.
Performance Variable?

- **Recognition Rate?**
  - Defined over a set of people, not per person.

- **Similarity score?**
  - Defined per person.
  - Linear models, ...
  - But, what does this tell us about actual performance?

- **Probability of being recognized at Rank 1?**
  - Defined per person.
  - Non-linear modeling problem.

- **Probability of being correctly verified at given FAR?**
  - Defined per person.
  - Non-linear modeling problem.
Statistical Modeling Overview

Sampled Normalized Similarity Scores

Covariates
- Algorithm
- Age
- Race
- Gender
- Skin
- Glasses
- Facial Hair
- Makeup
- Bangs
- Expression
- Mouth
- Eyes

Linear Model (ANOVA)

Predict:
similarity scores from covariate combinations
Statistical Modeling Overview

**Covariates**
- Algorithm
- Age
- Race
- Gender
- Skin
- Glasses
- FacialHair
- Makeup
- Bangs
- Expression
- Mouth
- Eyes

**Sampled Normalized Similarity Scores**

**Linear Model (ANOVA)**
- Predict: similarity scores from covariate combinations

**Generalized Linear Model**
- Predict: probability of correct recognition from covariate combinations

**Sampled Recognition Ranks**
Linear Model - Similarity (Distance)

\[ Y_i = \text{Similarity (Distance) metric for image pair } i. \]

\[ X_i = \text{Algorithm & Human covariate factors for image pair } i. \]

\[ \beta = \text{Parameters quantifying factor effects.} \]

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \varepsilon_i \]

with \( \varepsilon_i \sim \text{iid Normal}(0, \sigma^2) \)
Generalized Linear Model
Pr(correct rank one recognition)

Y_i = Was the ith image pair matched at rank 1?
    (i.e. Y_i = 1 if R_i = 1 and otherwise Y_i = 0)
X_i = Algorithm & Human covariate factors for
     image pair i.
β  = Parameters quantifying factor effects.

\[ g(\mu_{Y_i|X_i}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \epsilon_i \]
\[ Y_i | X_i \sim f(\mu_{Y_i|X_i}) \text{ independently} \]

Now: \[ g(z) = \log \left( \frac{z}{1-z} \right), \ f(\mu_{Y_i|X_i}) = \text{Bernoulli}(\mu_{Y_i|X_i}) \]
What Do Models Tell Us?
PCA Algorithm Example.

Look at age holding all other covariates fixed.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Base</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Young</td>
<td>Old</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>Skin</td>
<td>Clear</td>
<td>Clear</td>
</tr>
<tr>
<td>Bangs</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Expression</td>
<td>Neutral</td>
<td>Neutral</td>
</tr>
<tr>
<td>Eyes</td>
<td>Open</td>
<td>Open</td>
</tr>
<tr>
<td>Facial Hair</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Makeup</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mouth</td>
<td>Closed</td>
<td>Closed</td>
</tr>
<tr>
<td>Glasses</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Similarity Scores - LM
- 13.0% Increase in similarity
- p-value < 0.0001
- Older is easier.

Pr(rank-one) - GLM
- Pr(crk=1) = 0.916 Base
- Pr(crk=1) = 0.951 Old
- p-value = 0.009
- Older is easier.
What Do Models Tell Us?
PCA Algorithm Example.

Look at gender holding all other covariates fixed.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Base</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Young</td>
<td>Young</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>Skin</td>
<td>Clear</td>
<td>Clear</td>
</tr>
<tr>
<td>Bangs</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Expression</td>
<td>Neutral</td>
<td>Neutral</td>
</tr>
<tr>
<td>Eyes</td>
<td>Open</td>
<td>Open</td>
</tr>
<tr>
<td>Facial Hair</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Makeup</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mouth</td>
<td>Closed</td>
<td>Closed</td>
</tr>
<tr>
<td>Glasses</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Similarity Scores - LM
- 1.7% decrease in similarity
- p-value < 0.33
- Gender is not significant.

Pr(rank-one) - GLM
- Pr(crk=1) = 0.915 Base
- Pr(crk=1) = 0.884 Female
- p-value = 0.0925
- Gender is not significant
Model Validation & p-values

Table 1: ANOVA results for the linear model. ‘B’=‘both images’, ‘O’=‘Other’, ‘Ch’=‘changes from one image to the other’, and ‘:’ indicates an interaction.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est</th>
<th>S.E.</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-8.44</td>
<td>0.08</td>
<td>-107.76</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>IIDC</td>
<td>5.48</td>
<td>0.11</td>
<td>49.46</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>EBMGM</td>
<td>3.54</td>
<td>0.11</td>
<td>31.98</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Old</td>
<td>-0.57</td>
<td>0.08</td>
<td>-7.09</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Female</td>
<td>0.18</td>
<td>0.09</td>
<td>2.14</td>
<td>0.0324</td>
</tr>
<tr>
<td>Afr. American</td>
<td>-0.19</td>
<td>0.11</td>
<td>-1.76</td>
<td>0.0790</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.64</td>
<td>0.10</td>
<td>-6.43</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>O Race</td>
<td>-0.07</td>
<td>0.12</td>
<td>-0.59</td>
<td>0.5534</td>
</tr>
<tr>
<td>O Skin</td>
<td>-0.29</td>
<td>0.09</td>
<td>-3.08</td>
<td>0.0021</td>
</tr>
<tr>
<td>B Bangs</td>
<td>-0.82</td>
<td>0.08</td>
<td>-9.74</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Bangs Ch</td>
<td>-1.08</td>
<td>0.19</td>
<td>-5.63</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B O Expression</td>
<td>0.65</td>
<td>0.15</td>
<td>4.39</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Expression Ch</td>
<td>1.63</td>
<td>0.08</td>
<td>19.94</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B Eyes Not Open</td>
<td>-1.66</td>
<td>0.32</td>
<td>-5.22</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Eyes Ch</td>
<td>1.56</td>
<td>0.11</td>
<td>13.79</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B Facial Hair</td>
<td>0.25</td>
<td>0.10</td>
<td>2.40</td>
<td>0.0164</td>
</tr>
<tr>
<td>Facial Hair Ch</td>
<td>-0.75</td>
<td>0.32</td>
<td>-2.34</td>
<td>0.0191</td>
</tr>
<tr>
<td>B Glasses</td>
<td>-2.43</td>
<td>0.13</td>
<td>-18.14</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B Makeup</td>
<td>-0.23</td>
<td>0.11</td>
<td>-2.02</td>
<td>0.0439</td>
</tr>
<tr>
<td>Makeup Ch</td>
<td>0.32</td>
<td>0.26</td>
<td>1.23</td>
<td>0.2179</td>
</tr>
<tr>
<td>B O Face</td>
<td>0.14</td>
<td>0.11</td>
<td>2.21</td>
<td>0.0268</td>
</tr>
<tr>
<td>Mouth</td>
<td>1.17</td>
<td>0.10</td>
<td>11.7</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>IIDC Ch</td>
<td>0.37</td>
<td>0.11</td>
<td>3.22</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Table 2: Summary of generalized linear model results.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>df</th>
<th>ΔDeviance</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>Note 1</td>
<td></td>
</tr>
<tr>
<td>Algorithm</td>
<td>2</td>
<td>Note 2</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>5.73</td>
<td>0.0167</td>
</tr>
<tr>
<td>Bangs</td>
<td>2</td>
<td>63.99</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Facial Hair</td>
<td>2</td>
<td>11.12</td>
<td>0.0039</td>
</tr>
<tr>
<td>Mouth</td>
<td>2</td>
<td>76.50</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Race &amp; Alg. : Race</td>
<td>9</td>
<td>46.48</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Skin &amp; Alg. : Skin</td>
<td>3</td>
<td>24.00</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Expr. &amp; Alg. : Expr.</td>
<td>6</td>
<td>54.64</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Eyes &amp; Alg. : Eyes</td>
<td>6</td>
<td>131.87</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Glasses &amp; Alg. : Glasses</td>
<td>3</td>
<td>8.15</td>
<td>0.0430</td>
</tr>
<tr>
<td>Gender &amp; Alg. : Gender</td>
<td>3</td>
<td>9.55</td>
<td>0.0228</td>
</tr>
</tbody>
</table>

Note 1 The null model deviance is 4,266.9 on 6,425 df. The model using all terms given above has residual deviance of 3,676.9 on 6,386 df—highly significant.

Note 2 The factor indicating algorithm has many significant interactions in this model and is highly significant. In a table organized to show subject covariate effects, an analogous test for algorithm would be distracting.

• Don’t try to read this…
• Standards for evaluating and reporting results important.
Age: Young vs. Old

Subject
Old

HARDER EASIER

-0.02 0.0 0.02

Change to Baseline Predicted Pr(crk=1)

EBGM
IIDC
PCA
Eyes: Open vs. Closed

Eyes Closed

Change to Baseline Predicted $Pr(\text{crk}=1)$

EBGM
IIDC
PCA
Verification Performance

Covariates
- Age
- Gender
- Bangs
- Facial Hair
- Eyes

Sampled verification outcomes at different false alarm rates

Generalized Linear Mixed effect Model (GLMM)

Predict: probability of correct verification from covariate and false alarm rate combinations
Verification Outcomes at Fixed False Alarm Rate $\alpha$

First Image of Each Person

Second Image of Each Person

Two Images per Subject Example
50 x 50 Similarity Matrix
Verification Outcomes at Fixed False Alarm Rate $\alpha$

1) Set FAR $\alpha$, e.g. $\alpha = 1/250$

Two Images per Subject Example
50 x 50 Similarity Matrix
Verification Outcomes at Fixed False Alarm Rate $\alpha$

1) Set FAR $\alpha$, 
\[ \text{e.g. } \alpha = 1/250 \]

2) Indicate people correctly verified at threshold corresponding to $\alpha$

Two Images per Subject 
Example 
50 x 50 Similarity Matrix
Verification Indicator Variable and FAR settings

- Our study - 1,072 x 1,072 similarity matrix.
  - 1,072 match scores,
  - 1,148,112 non-match scores.

Indicator Variable $Y$ for each subject for each FAR setting:
  - 1 verified
  - 0 otherwise

<table>
<thead>
<tr>
<th>Setting</th>
<th>FAR ($\alpha$)</th>
<th>Rate per 10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/10,000</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1/5,000</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1,2,500</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1/1,000</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>1/500</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>1/250</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>1/100</td>
<td>100</td>
</tr>
</tbody>
</table>
Linearity of Log Odds against Log FAR - FERET+PCA

\[ \ln \left( \frac{VR}{1 - VR} \right) \]

\[ \ln \left( \frac{p}{1 - p} \right), \ p = \text{verification probability} \]
Linearity of Log Odds against Log FAR - FRVT
Generalized Linear Mixed Model (GLMM)

Analysis is: *Mixed Effects Logistic Regression with Repeated Measures on People.*

- Let A and B be 2 factors that might influence algorithm performance. For example, age and gender.
  - Example factor settings A=a and B=b.
- Let j index the FAR setting, $\alpha_j$.
- $Y_{pabj}$ is
  - 1 if Person p is verified correctly,
  - 0 otherwise.
- $Y_{pabj}$ depends on:
  - person p,
  - factors A and B, and
  - false alarm rate $\alpha_j$. 

Ross Beveridge, Biometric Quality Workshop, March 9, 2006
GLMM Model Continued …

\[ Y_{pabj} \] is Bernoulli R.V. with success probability \( p_{pabj} \)

\[
\log \left( \frac{p_{pabj}}{1 - p_{pabj}} \right) = \mu + A_a + B_b + \gamma_j \log(\alpha_j) + A_a \gamma_{aj} \log(\alpha_j) + \pi_p
\]

- \( \mu \) = grand mean
- \( A_a \) = effect of setting \( a \) of factor \( A \)
- \( B_b \) = effect of setting \( b \) of factor \( B \)
- \( \gamma_j \log(\alpha_j) \) = log linear effect of \( \alpha_j \)
- \( \gamma_{aj} A_a \log(\alpha_j) \) = interaction effect
- \( \pi_p \) = subject id. random effect (next page)
Subject Variation - The Mixed in Generalized Linear **Mixed** effect Model

\[
\begin{bmatrix}
\pi_1, \ldots, \pi_{1,072}
\end{bmatrix}^T \sim \text{Multivariate Normal where}
\]
\[
E(\pi_p) = 0, \quad \text{Var} \quad \pi_p = \sigma^2_p,
\]
\[
\text{Cor}(y_{pab\alpha}, y_{p'a'b'\alpha'}) = \begin{cases} 
\phi & \text{if} \quad p = p' \\
0 & \text{if} \quad p \neq p'
\end{cases}
\]

This means:

The outcomes, i.e. verification success/failure, are uncorrelated when testing different people but correlated when testing the same person under different configurations.
Random Effects are Important

GLMM vs. GLM

• Some people are harder to recognize than others.
• But, we don’t care who specifically is hard or easy.

Removing the “noise” of random effects helps reveal other significant effects of interest.
Marginal Verification Rates - Age

Verification Frequency

False Accepts per 10,000

old
young

Ross Beveridge, Biometric Quality Workshop, March 9, 2006
Results of the Model - Age

![Graph showing fitted verification probability vs. false accepts per 10,000. The graph compares 'old' and 'young' categories.](image)

- **Fitted Verification Probability**
- **False Accepts per 10,000**

Ross Beveridge, Biometric Quality Workshop, March 9, 2006
Marginal Verification Rates - Bangs

![Graph showing the relationship between verification frequency and false accepts per 10,000 with two lines: one for both bangs and another for neither bangs.](image)

- **Verification Frequency**
- **False Accepts per 10,000**
- **Both bangs**
- **Neither bangs**
Results of the Model - Bangs
Step Back: Why use Linear Models and Generalized Linear Models

Start with a set of factors - covariates

These may be …

Properties of the subject: age, etc.

Properties of the scene: lighting, etc.

Properties of the image:

Focus

Resolution

Contrast

…
Step Back: Why use Linear Models and Generalized Linear Models

A Descriptive Function, Probability of success given covariate values

May be a Quality Measure

\[ P(\text{success} \mid F_1, F_2, F_3, \ldots, F_k) \]
Thank You
LM with Three Algorithms

Subject Female  
Both Glasses  
Eyes Change  
Both Eyes Not Open  
Expression Changes  
Both Other Expression  
Subject Other Skin  
Subject Other Race  
Subject Asian  
Subject Black  
Mouth Changes  
Both Other Mouth  
Facial Hair Changes  
Both Facial Hair  
Bangs Change  
Both Bangs  
Subject Old  
Makeup Changes  
Both Makeup

Change to Baseline Standardized Distance

Page 42