Applying Machine Learning to Predict and Explain Primate Consortship
Josh King, Vayu Kishore, Filippo Ranalli  {jking9,vayu,franalli}@stanford.edu

We investigate the reasons for success and failure of mating between wild yellow baboon pairs. Our analysis applies classification methods to examine whether successful consortships can be predicted, and whether certain behavioral or genetic features are especially relevant in determining consortship.

Data

Our input dataset, from Tung et al (2012) is a set of genetic and behavioral features between potentially mating pairs, and a label indicating whether consortship occurred.
- 12,000 observations, 115 females, 121 males
- 1648 consorts, 10493 non-consorts

Additional preprocessing applied to standardize features and remove points that had conflicting labels. We also tried PCA whitening and data augmentation but found that they did not improve results.

Features

• Genetic/Biological: two measure of genetic diversity of each of the pair, estimated genetic distance of the pair, age, conceptiveness of female.
• Social/Behavioral: male rank, males/females present during consortship.
• Transformations: non-linear transformations of genetic distance rank, and age, as well as indices derived from the combination of genetic diversity of the pair.
• Graph-Based: PageRank, HITS.

Models

• Gaussian SVM
  \[ f(w, b) = \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i (w \cdot \phi(x_i) + b)) + \lambda ||w||^2 \]
• AdaBoosting with Decision Stumps
  \[ E_T = \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i (\sum_{t=1}^{T} \alpha_t h_t(x_i))) \]
• Random Forest
• Edge Prediction

For all models, Classes were weighted to address imbalanced dataset.

Results

<table>
<thead>
<tr>
<th>Features</th>
<th>Error</th>
<th>Precision</th>
<th>Recall</th>
<th>Fischer AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdaBoosting</td>
<td>0.394</td>
<td>0.210</td>
<td>0.563</td>
<td>0.277</td>
</tr>
<tr>
<td>Gaussian SVM</td>
<td>0.375</td>
<td>0.221</td>
<td>0.652</td>
<td>0.318</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.368</td>
<td>0.214</td>
<td>0.656</td>
<td>0.323</td>
</tr>
</tbody>
</table>

Future Work

• Model groups separately.
• Continue to investigate graphical methods and other ways to augment features for existing dataset.
• Find or generate a larger dataset with additional features.

References