Algorithms for Combining Classifiers with Applications to Face Recognition

Yehuda Vardi, Cun-Hui Zhang and Ofer Melnik
Rutgers University
Classifier Combination

Goals:
- Improve performance over constituent classifiers.
- Maximize information use.

Challenges:
- Intelligent combination that exploits complementary information, while not amplifying errors.
Combination as scoring of rankings

- Score classes using multiple rankings

<table>
<thead>
<tr>
<th>Class</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>3</td>
</tr>
<tr>
<td>Dick</td>
<td>1</td>
</tr>
<tr>
<td>Harry</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>2</td>
</tr>
<tr>
<td>Dick</td>
<td>1</td>
</tr>
<tr>
<td>Harry</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>6.3</td>
</tr>
<tr>
<td>Dick</td>
<td>10</td>
</tr>
<tr>
<td>Harry</td>
<td>5.1</td>
</tr>
</tbody>
</table>
Optimal Score Functions

- The optimal score function scores each rank combination by its likelihood when it is the correct class:

\[ f_\theta(r^{(1)}, \ldots, r^{(J)}) \sim P \left\{ R^{(j)}(k) = r^{(j)}(k), j \leq J, k \leq K \mid q \sim t_\theta \right\} \]

- Optimal score function requires exponential amounts of data for estimation.

- Can only be estimated for small numbers of classes (not biometrics) and small numbers of classifiers.
State-of-the-art Rank Combination
Score Functions

- **Fixed**: Borda Count, Highest Rank
- **Trained**: Logistic Regression
Borda Count

- The score of a class is the (negative) sum of its ranks across all classifiers.

\[ S(\theta) \equiv f_\theta(r^{(1)}, \ldots, r^{(J)}) = \sum_{j=1}^{J} -r^{(j)}(\theta) \]

- A democratic approach, treating all classifiers and ranks equally.
Highest Rank

- The score of a class is equal to the best rank it received.

\[ S(\theta) \equiv f_\theta(r^{(1)}, \ldots, r^{(J)}) = - \min \left( r^{(1)}(\theta), \ldots, r^{(J)}(\theta) \right) \]

- Assumes the classifier that gave a class the best rank is the most confident.
Logistic Regression

Generalization of Borda Count, score is a weighted sum of ranks.

\[ S(\theta) \equiv f_\theta (r^{(1)}, \ldots, r^{(J)}) = - \sum_{j=1}^{J} w_j r^{(J)} (\theta) \]

The weights reflect preference for individual classifiers. They are estimated using a separate combination training set.
Mixed Group Ranks (MGR)

- A generalization of Logistic Regression and Highest Rank

\[
S \left( r^{(1)}(\theta) = x_1, \ldots, r^{(J)}(\theta) = x_J \right) = \sum_{A \subseteq \{1, \ldots, J\}} -w_A \min \{x_j : j \in A\}
\]

- Simple interpretation for two classifiers

\[
S(x_1, x_2) = -w_1 x_1 - w_2 x_2 - w_{12} \min (x_1, x_2)
\]
MGR Structure

- Is linear for every permutation of rank orderings:

\[ x_{q_1} \leq x_{q_2} \leq \ldots \leq x_{q_J} \]

\[
S_{x_{q_1} \leq x_{q_2} \leq \ldots \leq x_{q_J}} = - \sum_{A \cup q_1 | A \subseteq \{1, \ldots, J\}} w_A x_{q_1} - \sum_{A \cup q_2 | A \subseteq \{1, \ldots, J\} - q_1} w_A x_{q_2} - \ldots - w_{q_J} x_{q_J}
\]

- Lower ranks get higher weights (confidence in lower ranks).
FERET Experiments Datasets

Five datasets

- First dataset used as gallery for comparison.
- Four other datasets (fafb, dup I, dup II, fafc) used for combination experiments. Pick one as combination training set, others for testing.
FERET Experiments Classifiers

- University of Southern California submission (USC97)- elastic bunch graphing approach.
- University of Maryland submission (UMD97)- discriminant analysis of Eigenfaces.
- ANM- baseline algorithm using Mahalanobis angular distance of Eigenfaces.
MGR 2 Classifier Experiments

![Graphs showing cumulative performance for different training and testing configurations]

- Train fab / Test dup II
- Train dup II / Test fabc
- Train fabc / Test fabf
- Train fabf / Test dup II
- Train dup I / Test fabc
- Train fabc / Test dup I
- Train fabf / Test dup I
- Train fabf / Test fabc
Combining with Mixed Group Ranks

◆ By generalizing the linearly democratic approach of Logistic Regression with the hard rules of Highest Rank, it can adaptively emphasize the recognizers exhibiting more accuracy while still incorporating all information.
MGR 3 Classifier Experiment
Experiments Summary

- MGR is a robust combination approach - significant performance increases are seen on training/testing combinations of easy/hard, hard/easy and hard/hard, surpassing the performance of Logistic Regression and Highest Rank in most cases.
Theoretical Basis of MGR
Properties of Score Functions

- Score functions are monotonic. Larger ranks mean lower scores.
  \[ \forall i, \; x_i \geq y_i \implies S(x) \leq S(y) \]

- The sublevel sets of score functions are convex
  \[ f(x) \geq f(y) \iff f(x) \geq f(\lambda x + (1 - \lambda)y) \]
So what are the good score functions?

- Score function with *confidence in lower ranks* (e.g. HR) and *preference* toward more accurate classifiers.
Summary

- Mixed Group Ranks is a successful and robust combination method appropriate for high-dimensional biometrics that generalizes the confidence in low ranks (e.g. HR) and preference of LR.
- Based on a theoretical framework forming the foundation for new combination approaches.
- Potential applications in single and multiple modalities of biometrics.