Introducing the Separability Matrix for ECOC coding

Miguel Angel Bautista, Sergio Escalera, Xavier Baró & Oriol Pujol
Outline

• Classification problems and the ECOC framework.
• Motivation.
• The Separability Matrix.
• An application of the Separability Matrix for coding ECOCs.
  • The Confusion-Separability Extension Coding.
• Experiments and Results.
• Conclusions and Future Work.
Introduction to the ECOC framework

• Classification tasks are a well known type of supervised learning problem. The goal is to classify an object among a certain number of possible categories.
• The ECOC framework has proven to be a powerful tool to deal with multi-class classification problems.
• This framework is composed of two different steps:
  • Coding: Decompose a given N-class problem into a set of n binary problems.
  • Decoding: Given a test sample s, determine its category.
Introduction to the ECOC framework

- At the coding step a decomposition of the \( N \)-class problem into \( n \) binary problems is build and represented into a matrix.

\[
M_{N \times n} \in \{-1, +1, 0\}
\]

- The columns of the matrix represent the binary problems.

- The rows of the matrix represent the codes of the \( N \) classes.
Introduction to the ECOC framework

- At the decoding step a new sample $s$ is classified by comparing the binary responses to the rows of $M$ by means of a decoding measure $\delta$.
- Different types of decoding based on the distance used (i.e. Hamming Decoding, Euclidean Decoding, etc.)

$$\arg \min_i \delta(x^s, y^i)$$
Motivation

• Standard predefined or random strategies may not be suitable for a given problem.
• Find an optimum coding matrix $M$ for a given problem was proved to be an $NP$-Complete problem [1].
• In [2] we show how reduced codes can perform has well as standard designs with far less number of dichotomizers.
• Those reduced codes can be extended in a problem-dependent way to benefit from error-correcting principles.

[1] On the learnability and design of Error Correcting Output Codes, K.Crammer & Y. Singer
The Separability Matrix

- The Separability matrix $S$ contains the pairwise distance $\delta$ between the codewords in $M$.
- With this matrix we can analyze the correction capability $\rho$ of $M$ since,

$$\rho = \frac{\min(S) - 1}{2}$$

- Standard ECOC designs show constant Separability matrices.
An application of the Separability Matrix for coding ECOCs

• In [2] we show how reduced codes can perform has well as standard designs with far less number of dichotomizers.

• Identify those classes that need an increment of distance in order to benefit from error-correcting principles.
The Confusion-Separability Extension coding (CSE Coding)

- Use the **Confusion matrix** over a validation subset to find the most confused classes.
- Use the **Separability matrix** to find the classes that need at increment of distance in order to benefit from error-correcting principles.
- Compute and **Extension Matrix** of a Binary Compact ECOC matrix which is focused on the classes that show both **high confusion and low separation**.
- Extend the coding matrix until a **maximum length of N dichotomies**.
The CSE Coding Algorithm

1. Find the most confused classes: \( \text{arg max}_{c_i, c_j} (C_{i,j} + C_{j,i}) \)
2. Compute and Extension matrix \( E \) which increments \( \delta(y^i, y^j) \geq q \)
3. Fill the empty Extension codes taking into account the confusion of \( c^i, c^j \) with the rest of the classes.
4. Update Confusion and Separability matrices.

\[
\begin{align*}
M_{N \times n} &= \begin{bmatrix}
6 & 0 & 0 & 0 & 0 \\
0 & 4 & 0 & 2 & 0 \\
0 & 0 & 3 & 0 & 2 \\
0 & 0 & 0 & 3 & 2 \\
0 & 0 & 0 & 0 & 5 \\
\end{bmatrix} \\
S_{N \times N} &= \begin{bmatrix}
\star & \diamond & \triangle & \square & \bigcirc \\
0 & 1 & 1 & 3 & 2 \\
1 & 0 & 2 & 2 & 3 \\
1 & 2 & 0 & 2 & 1 \\
3 & 2 & 2 & 0 & 1 \\
2 & 3 & 1 & 1 & 0 \\
\end{bmatrix} \\
E_{N \times k} &= \begin{bmatrix}
2^k & \kappa \geq N \\
\end{bmatrix} \\
q &= 3
\end{align*}
\]

（10）
Experiments:

• We tested the novel methodology on several public datasets from the UCI Machine Learning Repository.

<table>
<thead>
<tr>
<th>Problem</th>
<th>#Training samples</th>
<th>#Features</th>
<th>#Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dermathology</td>
<td>366</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>Ecoli</td>
<td>336</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Segmentation</td>
<td>2310</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Glass</td>
<td>214</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Vowel</td>
<td>990</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Yeast</td>
<td>1484</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

• In addition we perform experiments with 3 public Computer Vision problems.
  • The ARFace dataset with 20 classes.
  • The Traffic Sign dataset with 36 classes.
  • The MPEG dataset with 70 classes.
Experiments: Methods and Settings

• We compare the One vs. All and Dense Random with the CSE coding with
  \( q \in \{3, 5\} \)
• Classification results are the average over a Stratified 10 fold CV.
• We use the SVM-RBF and AdaBoost as our base classifier.
• An optimization process is carried out to tune the parameters of the SVMs.
• SVM-RBF classifiers have 2 parameters to optimize (C & \( \gamma \)).
Experiments and Results

• Results for UCI and Computer Vision experiments with SVM as the base classifier.

<table>
<thead>
<tr>
<th>Data set</th>
<th>One vs. All ECOC Perf.</th>
<th>One vs. All ECOC Classif.</th>
<th>CSE ECOC $q = 3$ Perf.</th>
<th>CSE ECOC $q = 3$ Classif.</th>
<th>CSE ECOC $q = 5$ Perf.</th>
<th>CSE ECOC $q = 5$ Classif.</th>
<th>Dense Random ECOC Perf.</th>
<th>Dense Random ECOC Classif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel</td>
<td>55.0±10.5</td>
<td>11</td>
<td>66.9±7.8</td>
<td>9.2</td>
<td>69.8±6.3</td>
<td>10.6</td>
<td>67.9±8.3</td>
<td>11</td>
</tr>
<tr>
<td>Yeast</td>
<td>41.0±7.3</td>
<td>10</td>
<td>54.7±11.8</td>
<td>5.7</td>
<td>53.0±9.3</td>
<td>9.5</td>
<td>54.9±6.4</td>
<td>10</td>
</tr>
<tr>
<td>Ecoli</td>
<td>78.9±3.5</td>
<td>8</td>
<td>76.4±4.4</td>
<td>7</td>
<td>78.6±3.9</td>
<td>7.4</td>
<td>72.1±2.7</td>
<td>8</td>
</tr>
<tr>
<td>Glass</td>
<td>51.6±10.2</td>
<td>7</td>
<td>55.5±7.6</td>
<td>6</td>
<td>52.7±8.4</td>
<td>3</td>
<td>42.8±11.02</td>
<td>7</td>
</tr>
<tr>
<td>Segment</td>
<td>97.3±0.7</td>
<td>7</td>
<td>96.9±0.8</td>
<td>6.6</td>
<td>96.6±1.0</td>
<td>6.2</td>
<td>96.6±1.3</td>
<td>7</td>
</tr>
<tr>
<td>Derma</td>
<td>97.1±1.2</td>
<td>6</td>
<td>97.1±0.9</td>
<td>5.2</td>
<td>95.9±1.2</td>
<td>3</td>
<td>95.7±0.8</td>
<td>6</td>
</tr>
<tr>
<td>Vehicle</td>
<td>80.1±4.0</td>
<td>4</td>
<td>81.1±3.5</td>
<td>3</td>
<td>70.6±3.4</td>
<td>3</td>
<td>81.1±3.6</td>
<td>4</td>
</tr>
<tr>
<td>MPEG7</td>
<td>83.2±5.1</td>
<td>70</td>
<td>88.5±4.5</td>
<td>15</td>
<td>89.6±4.9</td>
<td>20.4</td>
<td>90.0±6.4</td>
<td>70</td>
</tr>
<tr>
<td>ARFaces</td>
<td>76.0±7.22</td>
<td>50</td>
<td>80.7±5.2</td>
<td>13.8</td>
<td>84.6±5.3</td>
<td>20.2</td>
<td>85.0±6.3</td>
<td>50</td>
</tr>
<tr>
<td>Traffic</td>
<td>91.3±1.1</td>
<td>36</td>
<td>95.7±0.92</td>
<td>12.2</td>
<td>96.6±0.8</td>
<td>19</td>
<td>93.3±1.0</td>
<td>36</td>
</tr>
<tr>
<td>Rank &amp; #</td>
<td>3.0</td>
<td>20.8</td>
<td>2.2</td>
<td>8.8</td>
<td>2.3</td>
<td>10.3</td>
<td>2.5</td>
<td>20.8</td>
</tr>
</tbody>
</table>
Experiments and Results

- Results for UCI and Computer Vision experiments with AdaBoost as the base classifier.

<table>
<thead>
<tr>
<th>Data set</th>
<th>One vs. All</th>
<th>CSE ECOC $\rho = 3$</th>
<th>CSE ECOC $\rho = 5$</th>
<th>Dense Random ECOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel</td>
<td>40.6±1.3</td>
<td>11</td>
<td>44.7±0.8</td>
<td>9.6</td>
</tr>
<tr>
<td>Yeast</td>
<td>36.8±1.1</td>
<td>10</td>
<td>45.6±0.4</td>
<td>9.6</td>
</tr>
<tr>
<td>Ecoli</td>
<td>71.5±10.9</td>
<td>8</td>
<td>68.1±8.3</td>
<td>7.4</td>
</tr>
<tr>
<td>Glass</td>
<td>53.8±12.1</td>
<td>7</td>
<td>52.8±13.5</td>
<td>6</td>
</tr>
<tr>
<td>Segment</td>
<td>96.4±0.7</td>
<td>7</td>
<td>95.0±0.3</td>
<td>6.8</td>
</tr>
<tr>
<td>Derma</td>
<td>89.3±4.9</td>
<td>6</td>
<td>77.6±6.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Vehicle</td>
<td>73.6±1.3</td>
<td>4</td>
<td>72.7±1.9</td>
<td>4</td>
</tr>
<tr>
<td>MPEG7</td>
<td>54.4±7.2</td>
<td>70</td>
<td>65.5±9.5</td>
<td>15</td>
</tr>
<tr>
<td>ARFaces</td>
<td>36.3±7.2</td>
<td>50</td>
<td>53.8±5.2</td>
<td>13.8</td>
</tr>
<tr>
<td>Traffic</td>
<td>80.6±6.2</td>
<td>36</td>
<td>81.3±8.1</td>
<td>12.2</td>
</tr>
<tr>
<td>Rank &amp; #</td>
<td>2.6</td>
<td>20.8</td>
<td>2.4</td>
<td>9.16</td>
</tr>
</tbody>
</table>
Conclusions and Future Work

- The **Separability Matrix** is introduced as a novel tool to analyze and enhance ECOC coding designs.
- The **Extension Algorithm** proposed can be applied to any existing ECOC scheme.
- A new coding design based on the Separability matrix is introduced **obtaining significant performance improvements** over state-of-the-art ECOC designs.
- The proposed methodology **reduces the number of base classifiers** needed in comparison with state-of-the-art designs.
- A possible improvement will be **to optimize the initial Compact ECOC coding matrix**.
Thank you!