Coding and Decoding Design of ECOCs for Multi-class Pattern and Object Recognition

Author: Sergio Escalera Guerrero

Advisors: Dr. Oriol Pujol Vila and Dr. Petia Radeva
Introduction

Error-Correcting Output Codes (ECOC)
  - Coding
  - Decoding

Applications

Conclusions
Main families of classifiers

- Similarity Maximization Methods
- Probabilistic Methods
- Geometric Classifiers

Multi-class versus binary classification

- Discriminative classifiers are binary-defined by default
- Most discriminative multi-class classifiers are defined as a combination of binary problems

Combining strategies

- Multi-class as a combination of binary classifiers (dichotomizers)
  - One-versus-all
  - One-versus-one (pairwise)

- Error-Correcting Output Codes

Error-Correcting Output Codes

- $C_1$  $C_2$
- $C_3$  $C_4$

$C_1$, $C_2$, $C_3$, $C_4$
Error-Correcting Output Codes - Coding

one-versus-one

Training

\[ C_1 \quad C_2 \quad C_3 \quad C_4 \]

\[ h_1 \quad h_2 \quad h_3 \quad h_4 \quad h_5 \quad h_6 \]

\[ = 1 \quad = -1 \quad = 0 \]

Introduction

ECOC

ECOC coding

ECOC decoding

Applications

Conclusions
Error-Correcting Output Codes - Decoding

one-versus-one

Testing

Introduction  ECOC  ECOC coding  ECOC decoding  Applications  Conclusions
Error-Correcting Output Codes

Properties

- The information among dichotomizers is used jointly to make a classification decision

- If the minimum distance among codewords is \( d \), then, \((d-1)/2\) errors can be corrected at the decoding step \([\text{Dietterich95}]\)

- ECOC method reduces the bias and the variance of the learning algorithm \([\text{Kong95}]\)

---


ECOC - Shortcomings

- **ECOC Coding:**
  - Pre-defined ECOC designs
    - Do not use the knowledge of the problem-domain

- **ECOC Decoding:**
  - Ternary ECOC framework [Allwein02]
    - Decoding strategies should be readjusted

---

ECOC coding – Classical strategies

one-versus-all
[Nilsson65]

one-versus-one
[Hastie98]

dense random
[Allwein02]

sparse random
[Allwein02]


ECOC coding – Our proposal
Forest-ECOC

- **Motivation**
  - Use the knowledge of the *problem-domain* to design the ECOC matrix
  - Take advantage of the embedding of a *tree structure* in the ECOC matrix to embed a Forest of trees [Pujol06]

---

**Forest-ECOC**

**Step 1**
- Generate the best tree structure

**Step 2**
- Embed each internal node of the tree as a new dichotomizer

**Introduction**
- ECOC
- ECOC coding

**Applications**

**Conclusions**

---

- $c_1 \Rightarrow y_1$
- $c_2 \Rightarrow y_2$
- $c_3 \Rightarrow y_3$
- $c_4 \Rightarrow y_4$

$h_1, h_2, h_3, h_4, h_5, h_6$
Forest-ECOC

- UCI Machine Learning Repository classification

<table>
<thead>
<tr>
<th></th>
<th>UCI</th>
<th>JB</th>
<th>all pairs FLDA</th>
<th>Forest ECOC</th>
<th>Dense random ECOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>1.57</td>
<td>1.57</td>
<td></td>
<td>1.42</td>
<td>3.0</td>
</tr>
</tbody>
</table>

- Small codewords
- Information of tree nodes is shared among classes in the ECOC matrix
ECOC coding – Our proposal
Properties

- Problem-dependent extension of any initial coding (even empty)
  - It focuses on difficult classes
    - It increases the distance between difficult to classify classes while preserving the rest
  - A validation subset is used to increase generalization and prevent or delay overfitting
ECOC Optimizing Node Embedding

Coding (Finding a new dichotomy)

Step 1
- Find the empirical error on the training and validation subsets

Step 2
- Select the pair of classes with highest error from the joint confusion matrix

Step 3
- Complete the sets of classes minimizing the joint error
  (Sequential Forward Floating Search strategy)

Introduction
ECOC
ECOC coding
ECOC decoding
Applications
Conclusions
Coding (Embedding)

**Embedding**

- Embed the new dichotomy in the matrix

\[
M(r, i) = \begin{cases} 
0 & \text{if } c_r \not\in C_i \\
+1 & \text{if } c_r \in C_{i_1} \\
-1 & \text{if } c_r \in C_{i_2} 
\end{cases}
\]

- Update the dichotomy importance

\[
w_i = 0.5 \log \left( \frac{1 - e_i}{e_i} \right)
\]

**Weighting**

1.0  |  weights
---|---
original code | extended code

**Original Code**

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Extended Code**

<table>
<thead>
<tr>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Boundaries resulted after one iteration of training. (a) ECOC-ONE, (b) one-versus-one, (c) one-versus-all and, (d) and (e) two different matrices of Dense Random with the same minimal distance, respectively. Dark line corresponds to the real boundary and grey regions correspond to learning errors.
Error evolution using ECOC-ONE with FLDA for:

(a) **Glass** data set

(b) **Dermathology** data set
The length of the codeword is increased in the way that a better solution for the training data is obtained.

It can be applied to any initial coding matrix, yielding a small code length.
ECOC coding – Our proposal

Introduction

ECOC coding

Applications

Conclusions

Forest

ECOC

ECOC coding

ECOC ONE

Sub-class

ECOC
Sub-class ECOC

- Linear classifier
- Adaboost
Step 1

- Find the next optimal tree node based on SFFS and Mutual Information

Step 2

- Test the node performance using a base classifier

Step 3

- If training error > epsilon:
  - Split the node data into subclasses maximizing inter-class cluster distance

Step 4

- Embed the new dichotomizers
Sub-class ECOC

Sub-class learning example

(a) FLDA
(b) Discrete Adaboost
(c) NMC

NMC

Linear SVM

RBF SVM

DECOC

Sub-class ECOC

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{11}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{12}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$h_1$  $h_2$  $h_3$

$C_{11}$  1  1  0
$C_{12}$  1  0  1
$C_2$     1 -1 -1
$C_3$     -1 0  0
Sub-class ECOC

- Sub-class ECOC splits the original set of classes into sub-classes until the base classifier is able to learn the training data.
- Useful when ECOC base classifier is not able to model the binary problems.
- It avoids the requirement of using complex classifiers.
ECOC decoding – Our proposal

- Attenuated Euclidean decoding
- Loss Weighted decoding
- Beta-density decoding
- Lapacian decoding
ECOC decoding – Classical strategies

Hamming decoding [Nilsson65]
\[ HD(x, y_i) = \sum_{j=1}^{n} (1 - \text{sign}(x^j \cdot y_i^j))/2 \]

Inverse Hamming decoding [Windeatt03]
\[ IHD(x, y_i) = \max(\Delta^{-1} D^T), \Delta(i_1, i_2) = HD(y_{i1}, y_{i2}) \]

Euclidean decoding [Hastie98]
\[ ED(x, y_i) = \sqrt{\sum_{j=1}^{n} (x^j - y_i^j)^2} \]

Loss-based decoding [Allwein02]
\[ LB(\rho, y_i) = \sum_{j=1}^{n} L(y_i^j \cdot f^j(\rho)) \]
\[ L(\theta) = -\theta , \quad L(\theta) = e^{-\theta} \]

Probabilistic-based decoding [Passerini04]
\[ PD(y_i, F) = -\log \left( \prod_{j \in [1, \ldots, n]: M(i,j) \neq 0} P(x^j = M(i, j)|f^j) + K \right) \]
\[ P(x^j = y_i^j | f^j) = \frac{1}{1 + e^{y_i^j(v^j f^j + \omega^j)}} \]

Definition 1: **Decoding bias** is the value introduced by the comparison of two codewords on positions containing the zero symbol (being the magnitude of the value proportional to the number of zero positions).

Definition 2: A **dynamic range bias** corresponds to the difference among the ranges of values associated to the decoding process of each codeword.
**Definition 3:** A general decoding decomposition to represent decoding strategies is defined as follows:

\[ d = \sum_{k \in I_b} b_k + \sum_{i \in I_a} a_i + \sum_{j \in I_e} e_j \]

- \( b \) – Value of a zero comparison
- \( a \) – Value of a matching
- \( e \) – Value of a failure
- \( I_b, I_a, I_e \) - Index

**Hypothesis I:** The bias induced by a zero position applying a particular decoding strategy should be zero (\( b=0 \))

**Hypothesis II:** The dynamic range should be constant for all the codewords
# Taxonomy

<table>
<thead>
<tr>
<th>Different dynamic ranges</th>
<th>$b \neq 0$</th>
<th>$b = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type I</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Same dynamic ranges</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
</table>

\[
d = \sum_{k \in I_b} b_k + \sum_{i \in I_a} a_i + \sum_{j \in I_e} e_j
\]

$b$ – value of a zero comparison, $a$ – value of a matching, $e$ – value of a failure

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$b$</th>
<th>$a$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>1/2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IHD</td>
<td>$-\frac{1}{2} W_1 + \sum_{i=2}^{N} \frac{W_i z_i}{z_1}$</td>
<td>0</td>
<td>$-1 W_1 + \sum_{i=2}^{N} \frac{W_i \beta_i}{\beta_1}$</td>
</tr>
<tr>
<td>ED</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>LLB&lt;sub&gt;C&lt;/sub&gt;</td>
<td>0</td>
<td>$-</td>
<td>f(\rho)</td>
</tr>
<tr>
<td>LLB&lt;sub&gt;D&lt;/sub&gt;</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>ELB&lt;sub&gt;C&lt;/sub&gt;</td>
<td>1</td>
<td>$1/e</td>
<td>f(\rho)</td>
</tr>
<tr>
<td>ELB&lt;sub&gt;D&lt;/sub&gt;</td>
<td>1</td>
<td>$1/e$</td>
<td>$e$</td>
</tr>
<tr>
<td>PDC</td>
<td>0</td>
<td>$\log \frac{1}{1+e</td>
<td>f(\rho)</td>
</tr>
<tr>
<td>PD&lt;sub&gt;D&lt;/sub&gt;</td>
<td>0</td>
<td>$\log \frac{1}{1+e}$</td>
<td>$\log \frac{1}{1+1/e}$</td>
</tr>
</tbody>
</table>
ECOC decoding – Our proposal

Attenuated Eucliden decoding

Loss Weighted decoding

Beta-density decoding

Lapacian decoding

ECOC decoding
Attenuated Euclidean decoding

- **Motivation**: Avoid the zero bias

\[
AED(x, y_i) = \sqrt{\sum_{j=1}^{n} |y_i^j| \sum_j \frac{\|x^j\|}{(x^j - y_i^j)^2}}
\]

\[
AED(x, y) = \sum_{j \in I_e} e_j = \sum_{j \in I_e} 4 = 4\beta
\]

- Still the dynamic ranges differ
ECOC decoding – Our proposal

- Attenuated Eucliden decoding
- Loss Weighted decoding
- Beta-density decoding
- Lapacian decoding

ECOC decoding

Introduction   ECOC   ECOC coding   ECOC decoding   Applications   Conclusions
Laplacian decoding

**Motivation:** Avoid the zero bias and make the dynamic ranges comparable

\[ d = \sum_{k \in I_b} b_k + \sum_{i \in I_a} a_i + \sum_{j \in I_e} e_j \]

- \( b \) – Value of a zero comparison
- \( a \) – Value of a matching
- \( e \) – Value of a failure
- \( I_b, I_a, I_e \) - Index

Define a measure that counts the number of coincidences between the input codeword and the class codeword.

In order to get constant dynamic ranges, the measure is normalized by the total number of positions coded by \([-1, +1]\):

\[ d(x, y_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \]

- Number of matches
- Number of failures
The main drawback of this definition is that it is not robust when there is a small number of coded positions in one word.

We introduce a prior bias, known as the Laplace Correction:

\[
LAP(x, y_i) = \frac{\alpha_i + 1}{\alpha_i + \beta_i + K}
\]

where \( K \) is an integer value that codifies the number of classes considered by the classifier – two in this case.
ECOC decoding – Our proposal

- Attenuated Euclidean decoding
- Loss Weighted decoding
- Beta-density decoding
- Lapacian decoding

ECOC decoding
**Motivation**: introduce confidence to the previous Laplacian decoding approximation

- Based on PDF estimation between two codewords
- Model the accuracy and uncertainty based on a pessimistic score
- We use an extension of the continuous binomial distribution, the Beta-distribution:

\[
\psi_i(\nu, \alpha_i, \beta_i) = \frac{1}{K} \nu^{\alpha_i} (1 - \nu)^{\beta_i}
\]

- The class which achieves the highest score \( s \), defined as the pessimistic score, is assigned to the test codeword:

\[
s_i : \int_{\nu_i-s_i}^{\nu_i} \psi_i(\nu, \alpha_i, \beta_i) d\nu = u
\]

where \( u \) is a threshold parameter. We fixed \( u=1/3 \) to govern the uncertainty influence.
The approach correctly classifies the example
The confidence grows with the sharpness of the PDF
ECOC decoding – Our proposal

- ECOC decoding
- Attenuated Euclidean decoding
- Loss Weighted decoding
- Beta-density decoding
- Laplacian decoding
**Motivation**: decoding decomposition, defining a matrix that weights the decoding process to assure that the hypotheses are fulfilled

\[ d = M_W \cdot T \quad , \quad T \text{ decoding measure} \]

1. **Step 1**
   - Compute the matrix of hypothesis \( H \)

2. **Step 2**
   - Normalize \( H \) to obtain a matrix \( M_W \) of PDF-rows

3. **Step 3**
   - Use \( M_W \) to weight the decoding process in a Loss-based decoding

\[ d(\varphi, i) = \sum_{j=1}^{n} M_W(i, j)L(M(i, j) \cdot f(\varphi, j)) \]
### Summary

<table>
<thead>
<tr>
<th>Different dynamic ranges</th>
<th>$b \neq 0$</th>
<th>$b = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type III</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Same dynamic ranges</th>
<th>$b \neq 0$</th>
<th>$b = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type III</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
d = \sum_{k \in I_b} b_k + \sum_{i \in I_a} a_i + \sum_{j \in I_e} e_j
\]

$b$ – Value of a zero comparison  
$a$ – Value of a matching  
$e$ – Value of a failure  
$I_b, I_a, I_e$ - Index

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$b$</th>
<th>$a$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AED</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>$\beta - DEN$</td>
<td>0</td>
<td>$\log(\nu)$</td>
<td>$\log(1 - \nu)$</td>
</tr>
<tr>
<td>LLW$_C$</td>
<td>0</td>
<td>$-MW(., i)</td>
<td>f(\rho)</td>
</tr>
<tr>
<td>LLW$_D$</td>
<td>0</td>
<td>$-MW(., i)$</td>
<td>$MW(., j)$</td>
</tr>
<tr>
<td>ELW$_C$</td>
<td>0</td>
<td>$\frac{MW(., i)}{e</td>
<td>f(\rho)</td>
</tr>
<tr>
<td>ELW$_D$</td>
<td>0</td>
<td>$\frac{MW(., i)}{e}$</td>
<td>$MW(., j)e$</td>
</tr>
</tbody>
</table>
Decoding evaluation

Nemenyi test

Stratified ten-fold cross-validation and two-tailed t-test

HD ED IHD LB PD AED
LAP BDEN LW

50 runs Gentle Adaboost and Linear SVM

1-vs-1 1-vs-all dense sparse decoc ecoc-one

16 UCI data sets
Decoding evaluation

UCI data sets

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Gentle Adaboost</th>
<th>Linear SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type 0</td>
<td>Type I</td>
</tr>
<tr>
<td>Discrete</td>
<td>5.5000</td>
<td>4.9844</td>
</tr>
<tr>
<td>Continuous</td>
<td>3.0799</td>
<td>2.7839</td>
</tr>
</tbody>
</table>
Applications

- Mobile Mapping
- Chagas disease
- Aibo
- IVUS
**Mobile Mapping System**

[Introduction] [ECOC] [ECOC coding] [ECOC decoding] [Applications] [Conclusions]

---

**Baro04**

Mobile Mapping – Forest-ECOC

Speed
- KNN
- TD – Tangent Distance
- PCA-KNN
- FLDA – Fisher+PCA
- SVM – Linear SVM
- BR – Gentle & Haar-like
- JB – Joint Boosting
- BS – Boosting Sampling
- NB – Naive Bayes
- Adaboost
- Forest-ECOC

Circular

Triangular
Mobile Mapping – Forest-ECOC

Speed group

Introduction

ECOC

ECOC coding

ECOC decoding

Applications

Conclusions
Mobile Mapping – Sub-class ECOC

<table>
<thead>
<tr>
<th></th>
<th>one-versus-one</th>
<th>one-versus-all</th>
<th>dense</th>
<th>sparse</th>
<th>DECOC</th>
<th>Sub-class ECOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global rank</td>
<td>1.8</td>
<td>3.6</td>
<td>3.4</td>
<td>4.6</td>
<td>2.6</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Discrete AdaBoost

- FLDA
- Linear SVM

NMC
Mobile Mapping – Decodings

- Coding: one-versus-one, one-versus-all, dense random, sparse random, decoc, ecoc-one
- Decoding: HD, ED, IHD, LLB, ELB, PD, AED, LAP, BDEN, LLW, ELW.
- Validation: stratified ten-fold cross-validation and test for the confidence interval with a two-tailed t-test
Applications

- Mobile Mapping
- Chagas disease
- IVUS
- Aibo
Sony Aibo robot

Multi-class syntethic signs detection by means of Adaboost with a Cascade of weak classifiers

ECOC classification
Applications

- Mobile Mapping
- Chagas disease
- Applications
- Aibo
- IVUS
IVUS

- Lipidic
- Calcium
- Fibrosis
Features

- RF, Slope, and Texture-based features [Karla06]

Data set

- We used the RF signals and their reconstructed images from a set of 10 different patients with Left Descent Artery pullbacks acquired in Hospital “German Trias i Pujol” from Badalona, Spain.

Statistical significance of sub-class strategy using Friedman and Nemenyi tests

Applications

- Mobile Mapping
- Chagas disease
- IVUS
- Aibo

Introduction
ECOC
ECOC coding
ECOC decoding
Applications
Conclusions
Chagas disease

- QRS Slope features [Pueyo07]
- 107 individuals grouped based on their degree of coronary damage
- Simon Bolivar University (Venezuela)

Conclusions

- Problem-dependent methodology to deal with the ECOC coding step:
  - Problem-dependent ECOC approaches yield compact codewords and thus lead to fast and robust classification rate avoiding overfitting
  - Forest-ECOC and ECOC-ONE extend the coding process based on the ensemble performance
  - Sub-class ECOC enriches the problem of ECOC design from the point of the view of the data

- Zero-bias free methodology to deal with the ECOC decoding:
  - Common taxonomy defined for all existing decoding strategies
  - Novel decoding strategies free from the zero bias proper to the classical ternary codewords that significantly improve the ECOC performance
  - Pessimistic Beta-Density Distribution decoding that gives a prediction based on modelling accuracy and uncertainty
  - Loss-Weighted decoding that overperforms other decoding strategies due to a weighting matrix applicable to any existing decoding strategy

- Viability on real-life applications
Future work

- Correspondence ECOC design versus base classifier
- Continuous ECOC construction
  - Binary $\rightarrow$ ternary $\rightarrow$ continuous
- Faster alternatives to ECOC coding designs constructions
- ECOC Public-domain toolbox
Codings:

Decodings:

Applications:

And more than 15 conference papers…
Thank you!!

Author: Sergio Escalera Guerrero

Advisors: Dr. Oriol Pujol Vila and Dr. Petia Radeva

9/7/2008