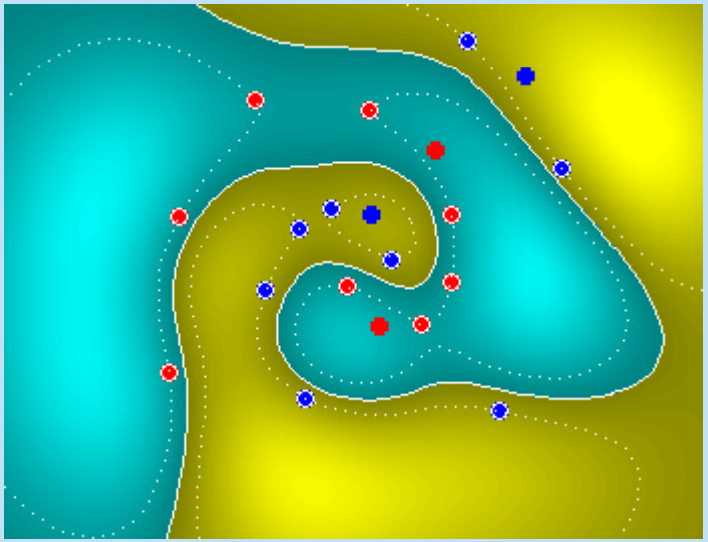
Problem-Dependent Designs for Error Correcting Output Codes: ECOC-ONE



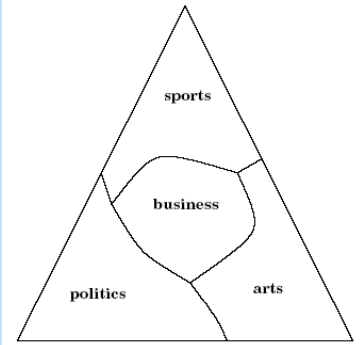
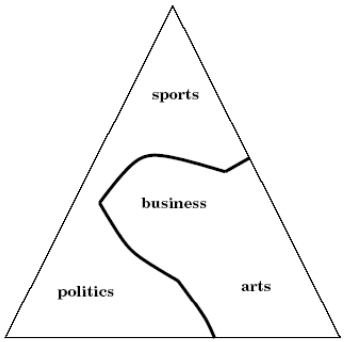
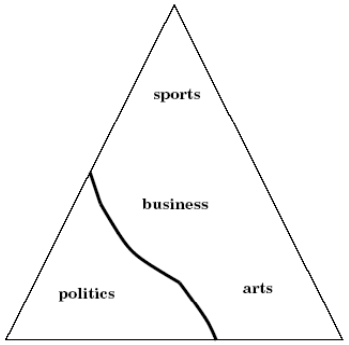
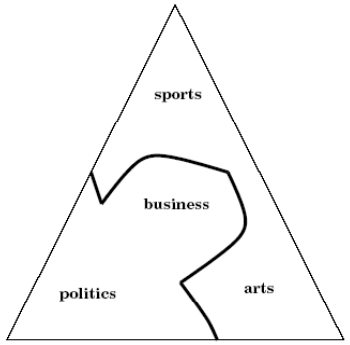
Sergio Escalera

Oriol Pujol

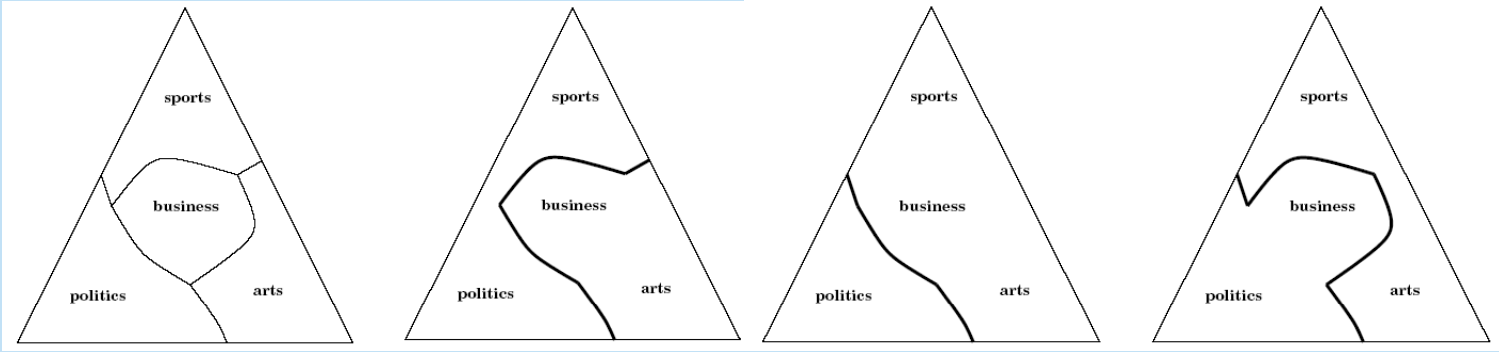
Petia Radeva

INDEX

1. Error Correcting Output Codes
2. Suboptimal node embedding in ECOC
3. Experiments and Results
4. Conclusions

Example	Classifier 1	Classifier 2	Classifier 3
			
C1= sports C2=business C3=politics C4=arts	1 -1 1 -1	1 1 -1 1	1 -1 -1 1

Example | Classifier 1 | Classifier 2 | Classifier 3

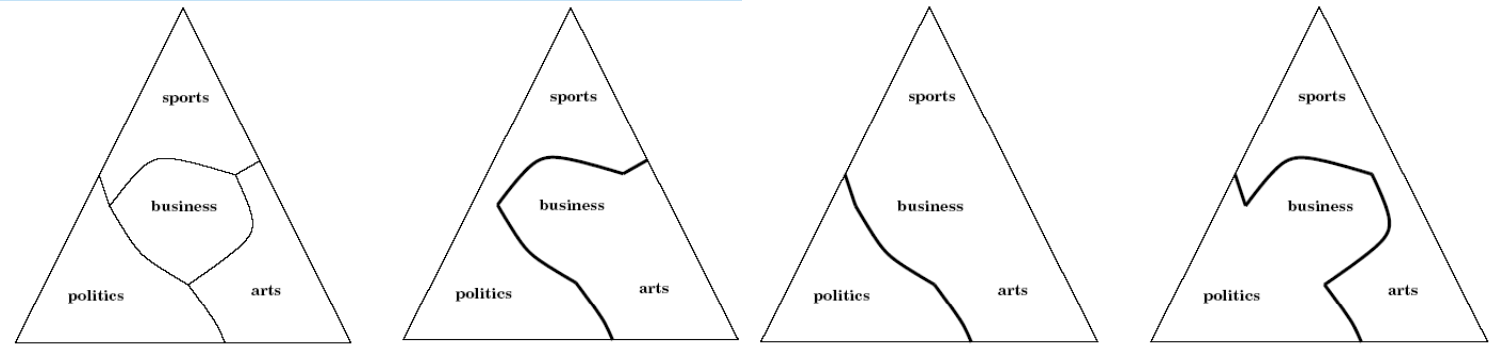


coding matrix

C1= sports	1	1	1
C2=business	-1	1	-1
C3=politics	1	-1	-1
C4=arts	-1	1	1

→ **code for class C4**

Example | Classifier 1 | Classifier 2 | Classifier 3



C1= sports
C2=business
C3=politics
C4=arts

coding matrix

1	1	1
-1	1	-1
1	-1	-1
-1	1	1

Decoding: Given a test sample, we obtain a code according to the output of each classifier and find the “closest” code.

-1 1 1

Standard strategies

Coding

One-vs-one

One-vs-all

Dense Random

Sparse Random

1 versus All

Code length: Nc

1	-1	-1
-1	1	-1
-1	-1	1

Random Dense ECOC

Code length: 10 log Nc

1	-1	1
-1	1	-1
1	-1	-1

Three symbol codes

1 versus 1: “All pairs”

Code length: Nc (Nc-1)/2

1	1	0
-1	0	1
0	-1	-1

Random Sparse ECOC

Code length: 15 log Nc

1	0	-1
-1	1	0
0	-1	1

Decoding

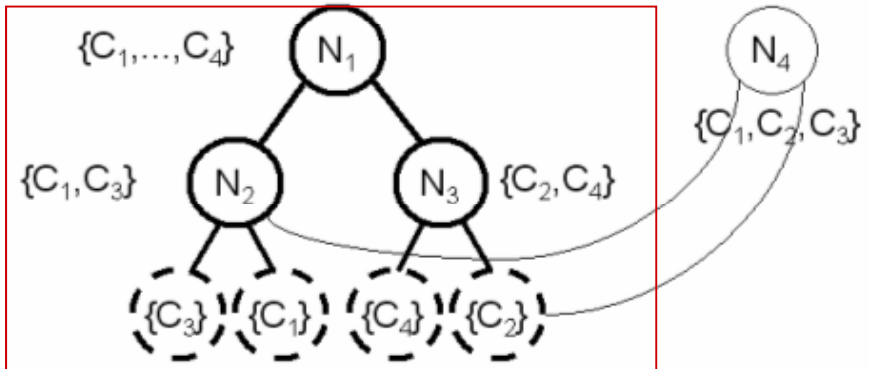
Hamming decoding

Euclidean decoding

Loss-based decoding

What we propose…

- Problem-dependent extension of any initial coding
  - Focus on difficult classes (increase the distance between difficult to classify classes while preserving the rest)



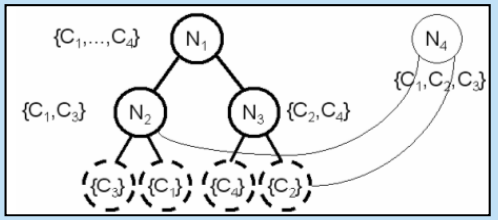
Optimal tree and first optimal node embedded.

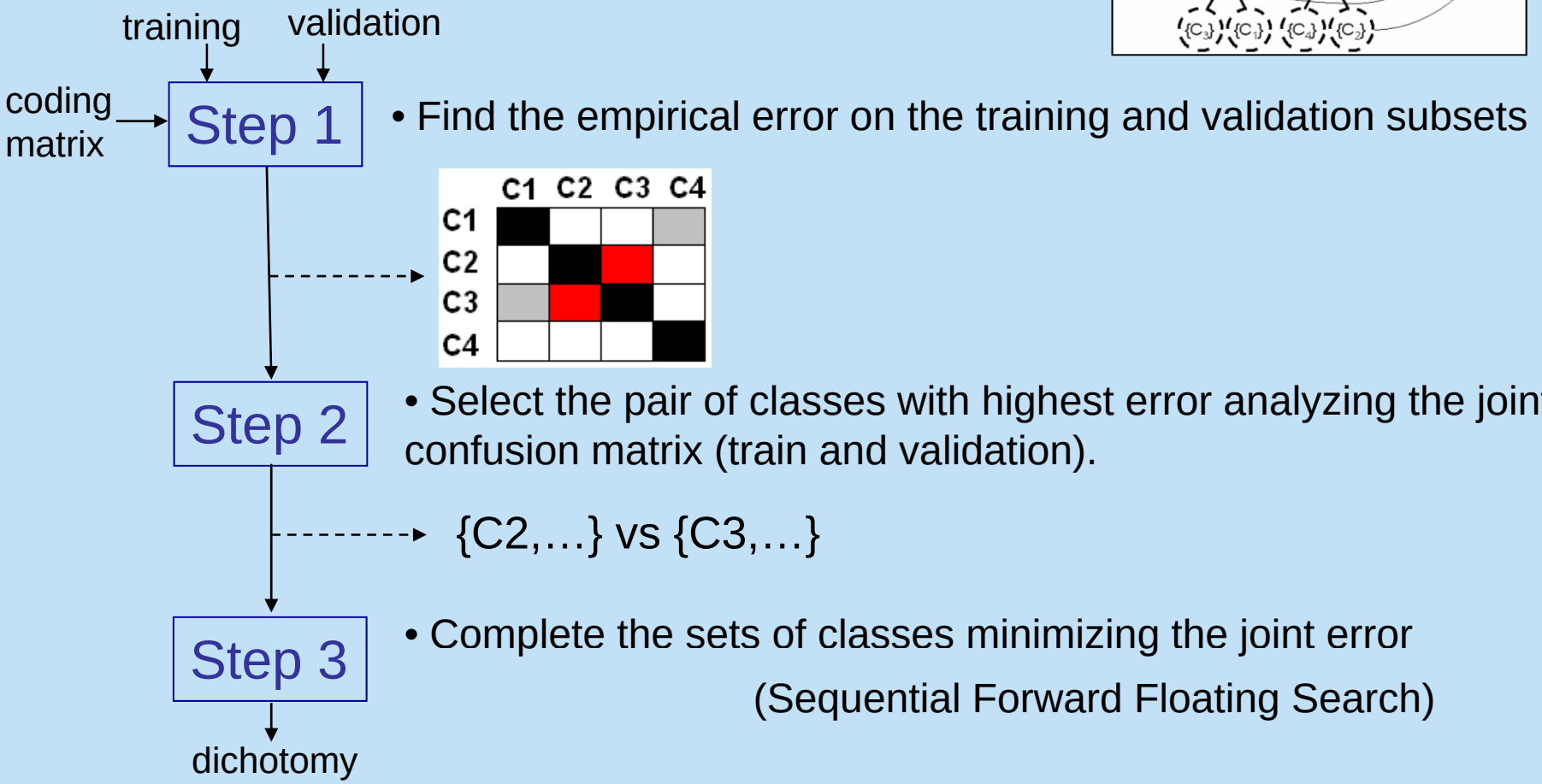
	N_1	N_2	N_3	N_4
C_1	1	-1	0	1
C_2	-1	0	-1	-1
C_3	1	1	0	1
C_4	-1	0	1	0

ECOC-ONE code matrix M for four dichotomies from the network.

- A validation subset guides the process to increase generalization and prevent or delay overfitting.

Coding (Finding a new dichotomy)





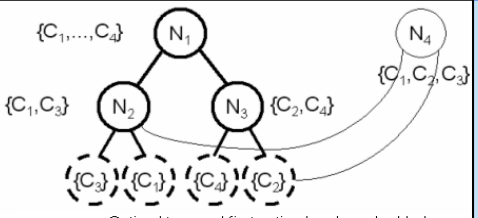
Coding

dichotomy {C2} vs {C3, C1}

Embedding

- Embed the new dichotomy in the matrix

$$M(r,i) = \begin{cases} 0 & if \quad c_r \notin C_i \\ +1 & if \quad c_r \in C_{i1} \\ -1 & if \quad c_r \in C_{i2} \end{cases}$$



Optimal tree and first optimal node embedded.

	N_1	N_2	N_3	N_4
C_1	1	-1	0	1
C_2	-1	0	-1	-1
C_3	1	1	0	1
C_4	-1	0	1	0

Weighting

- Update the dichotomy importance (weight)

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

1.0	weights
original code	extended code

Decoding (weighted attenuated Euclidean distance)

$$d = \sqrt{\sum_{i=1}^{n} |y_i| (x_i - y_i)^2 w_i}$$

UCI repository database

Use a Discriminant ECOC codification using classification score as split criterion instead of mutual information.

Problem	one-vs-one ECOC		one-vs-all ECOC		Dense random ECOC		ECOC-ONE	
	Hit	#D	Hit	#D	Hit	#D	Hit	#D
Dermathology	**96.30±0.61**	15	92.65±1.23	6	**95.26±0.82**	10	**95.17±0.74**	8.2
Ecoli	**78.05±1.46**	28	**77.10±1.19**	8	**77.65±1.33**	10	**78.15±1.84**	10
Glass	**67.93±1.66**	21	60.83±2.34	7	63.69±2.51	10	**67.03±1.63**	10
Segmentation	**97.01±0.72**	21	92.89±1.16	7	94.51±1.22	10	**96.23±1.52**	9.6
Vowel	**81.43±1.12**	55	73.33±1.40	11	74.50±1.96	10	**81.50±1.22**	10
Satimage	**86.23±0.79**	15	81.99±0.86	6	84.39±0.76	10	**85.47±1.00**	9.8
Yeast	**52.35±1.05**	45	**51.48±1.08**	10	**51.82±1.47**	10	**52.50±1.96**	10
Pendigits	**98.01±1.04**	45	93.98±2.56	10	**95.54±1.71**	10	**97.84±1.13**	10
Rank	**1.00**		1.88		1.63		**1.00**	

ECOC Strategies hits for UCI databases using Discrete Adaboost.

Note that one-vs-all with a voting policy for decoding corresponds exactly to the multiclass Adaboost (Adaboost.MH)

UCI repository database

Use a Discriminant ECOC codification using classification score as split criterion instead of mutual information.

Problem	one-vs-one ECOC		one-vs-all ECOC		Dense random ECOC		ECOC-ONE	
	Hit	#D	Hit	#D	Hit	#D	Hit	#D
Dermathology	**96.02±0.95**	15	**94.83±1.84**	6	**95.94±1.22**	10	**95.83±0.94**	8.7
Ecoli	**76.11±1.26**	28	63.97±1.51	8	72.94±1.37	10	**75.68±1.28**	10
Glass	**58.52±2.63**	21	49.73±2.45	7	54.13±2.73	10	**57.83±1.93**	10
Segmentation	**98.36±1.47**	21	94.36±1.13	7	93.83±1.43	10	**97.84±1.12**	9.2
Vowel	**73.18±1.15**	55	32.07±1.62	11	46.00±1.34	10	**69.14±3.01**	10
Satimage	87.43±0.80	15	85.85±1.08	6	84.03±1.49	10	**89.04±0.63**	10
Yeast	**55.31±1.47**	45	41.41±1.79	10	51.07±2.12	10	**52.58±1.73**	10
Pendigits	**98.53±1.03**	45	95.04±1.88	10	**96.44±1.12**	10	**98.43±0.99**	10
Rank	1.13		2.50		2.13		**1.00**	

ECOC Strategies hits for UCI databases using SVM.

UCI repository database

Extension of the one-versus-all strategy

Problem	one-vs-one ECOC		one-vs-all-ONE ECOC	
	Hit	#D	Hit	#D
Dermathology	**96.30±0.61**	15	**95.53±0.89**	8
Ecoli	**78.05±1.46**	28	**78.43±1.02**	10
Glass	**67.93±1.66**	21	**64.90±2.39**	9
Segmentation	**97.01±0.72**	21	**95.90±1.03**	9
Vowel	**81.43±1.12**	55	**79.34±1.40**	13
Satimage	**86.23±0.79**	15	**84.83±0.96**	8
Yeast	**52.35±1.05**	45	**53.52±0.89**	12
Pendigits	**98.01±1.04**	45	**96.88±2.01**	12

Conclusions

- We propose a coding framework to extend any ECOC.
  - It produces compact codewords (small length).
  - It is problem-dependent.
  - It focuses on difficult to discriminate classes increasing their code distance.
  - Its performance is comparable to the one-versus-one ECOC.

Open issues

- Embed other structures to capture knowledge from the domain.
- Develop efficient decoding strategies.

Thank you!