EVOLUTIONARY NEURAL NETWORK CLASSIFIERS TO PREDICT THE CLASSIFICATION OF R&D PERFORMANCE IN EU COUNTRIES

M. De la Paz-Marín, **P. Campoy-Muñoz**, C. Hervás-Martinez
ETEA, Business Administration Faculty, University of Cordoba
AYRNA Reserch Group, University of Córdoba
mpcampoy@etea.com

XVII SIGEF Congress "Methods for Decision Making in an uncertain environment"
6-8 Junio, Reus

Overview

- Purpose
  - Classify 25 UE countries' R&D performance during 2005-2008 and identify the most relevant variables
- Use
  - Help to monitor European strategies for R&D and innovation and some key features related to the EU innovation policy
- Methodology
  - Phase 1: k-means clustering
  - Phase 2: Multiclass classifiers (ESUNN, EPUNN)
  - Phase 3: Comparisson with other methods

The dataset

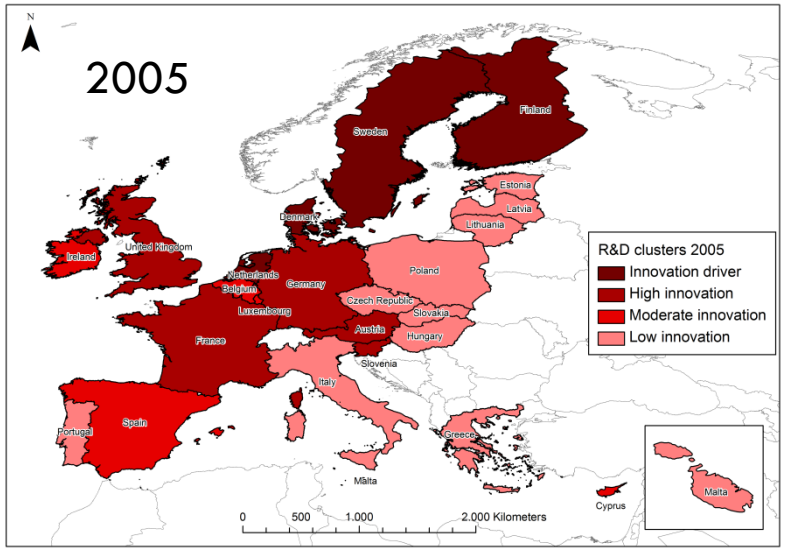
CODE	VARIABLE	UNIT
R&D enablers		
HUMLF	Human Resources in Science and Technology	% of labour force
RDPERS	R&D personnel (include all persons employed directly in R&D, plus persons supplying direct services)	% of labour force
GERDBU	Gross Domestic Expenditure on R&D: business sector	% of GDP
GERDGO	Gross Domestic Expenditure in R&D: government and Higher Education sectors	% of GDP
R&D results		
SPUBLI	Number of scientific publications	Per 1000 population
TPATRE	Worldwide patent applications filed through the Patent Cooperation Treaty procedure or with a national patent office (residents)	Per 1000 population
Education		
PHD06	Number of PhD graduates	Per 1000 population
TERTIT	Population having completed tertiary education	% of population aged 30-34
LLEARN	Lifelong learning	% of persons aged 18 to 64
Economy		
GDPGRO	Growth rate of GDP volume	% of change on previous year
RLPGH	Real labour productivity growth per hour worked	Real output per unit of labour input (measured by the total number of hours worked).
EMPLO	Total employment rate	% of persons aged 20 to 64 in employment
PATREV	License and patent revenues from abroad	% of GDP
TRADEM	Community trademarks	per billion GDP (in PPS€)
MHTEXP	Medium and High-technology exports	% of total manufactured exports

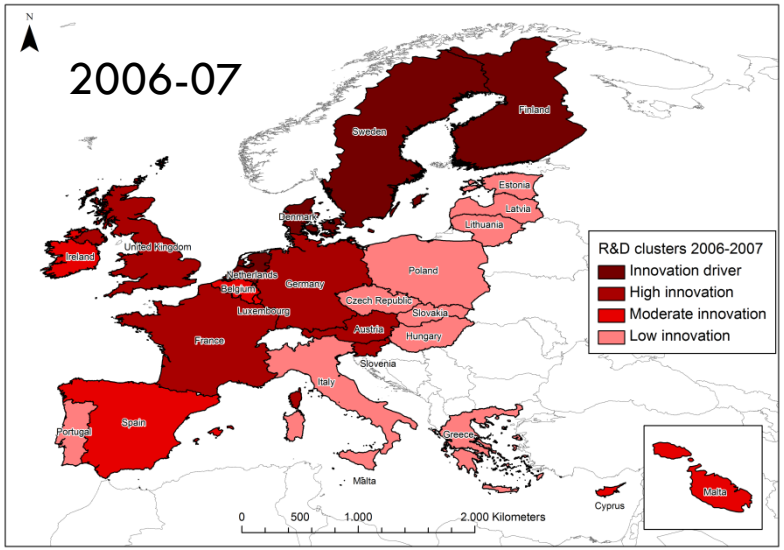
K-Means Clustering

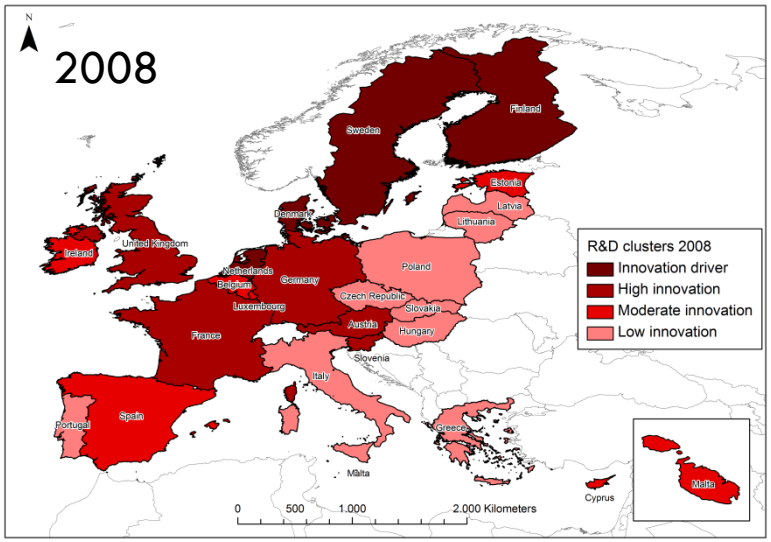
Variable	#pattern/ Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	100	24	16	40	20
HUMLF	31.39±6.08	30.09±6.03	40.52±6.08	27.70±6.08	33.04±6.11
RDPERS	0.99±0.48	1.00±0.48	1.59±0.49	0.66±048	1.17±0.47
GERDBU	0.94±0.72	0.78±0.69	2.00±0.73	0.40±0.72	1.38±0.71
GERDGO	0.56±0.21	0.41±0.21	0.89±0.21	0.45±0.21	0.70±0.21
SPUBLI	1.23±0.65	1.09±0.63	2.29±0.65	0.74±0.65	1.53±0.64
TPATRE	0.14±0.14	0.07±0.14	0.26±0.14	0.06±0.14	0.31±0.14
PHD06	0.17±0.09	0.10±0.09	0.25±0.09	0.14±0.09	0.24±0.09
TERTIT	31.09±10.29	38.20±10.39	41.58±10.40	22.79±10.29	30.76±10.56
LLEARN	15.90±6.71	12.76±6.54	27.26±6.80	11.71±6.71	18.94±6.92
GDPGRO	3.01±3.17	1.63±2.94	2.17±2.99	4.50±3.17	2.35±3.04
RLPGH	1.89±2.36	0.80±2.21	1.08±2.29	2.84±2.36	1.93±2.30
EMPLO	70.74±5.59	69.45±5.43	77.47±5.63	67.90±5.59	72.59±5.43
PATREV	0.39±0.51	0.55±0.53	1.00±0.52	0.12±0.51	0.27±0.52
TRADEM	4.59±3.22	7.65±3.34	5.38±3.28	2.39±3.22	4.69±3.24
MHTEXP	49.16±12.00	48.65±12.08	47.80±11.94	45.98±12.00	57.21±11.92

- **Moderate innovation** countries (Cluster 1)
- **Innovation-driven** countries (Cluster 2)
- **Low innovation** countries (Cluster 3)
- **High innovation** countries (Cluster 4)

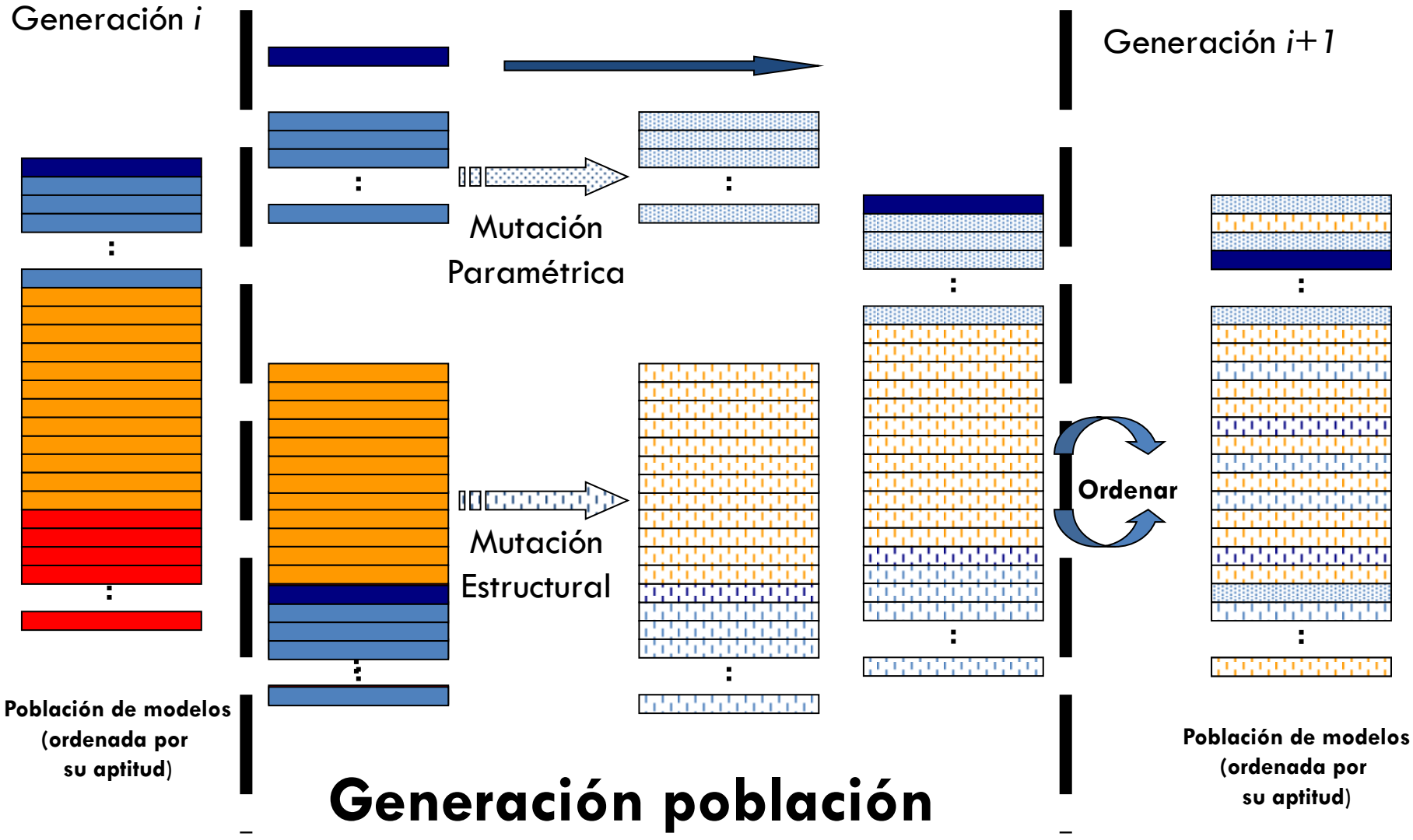
Mapping the clusters







Evolutionary algorithms: ESSUN and EPUNN



Experimental study

- Compared methodologies
  - Linear statistical model: SLDA
  - Neural network: MLP
  - Trees Clasifier: C4.5, LMT
  - Ensemble: AdaBoost 100
  - Logistic regression: SLogistic, Mlogistic
  - Kernel function: SVM,
  - Bayes network: Naïve Bayes

- Metrics
  - CC*R*→ performance in the whole dataset
  - *MS* →performance in each class
  - *K* →degree of association between pattern dsitributions in clases bedofe and after classification

Obtained results

Method	CCR_G(%)	MS_G(%)	K_G	#conn.
Best ESUNN	**100**	**100**	**1**	20
Best EPUNN	**100**	**100**	**1**	26
Best MLP	**100**	**100**	**1**	44
SLDA	96	80	0.94	24
C4.5	84	77	0.78	**13**
AdaBoost100	52	0	0.31	24
LMT	**100**	**100**	**1**	27
NaiveBayes	92	80	0.89	124
SLogistic	**100**	**100**	**1**	27
MLogistic	**100**	**100**	**1**	48
SVM	**100**	**100**	**1**	96

Method	CCR_G(%)	MS_G(%)	K_G	# conn.
	Mean±SD	Mean±SD	Mean±SD	Mean±SD
ESUNN	98.00±2.73	92.33±9.60	0.97±0.05	**26.57±3.47**
EPUNN	94.00±4.79	84.55±11.13	0.92±0.07	29.93±4.66
MLP	**100.00±0.00**	**100.00±0.00**	**1.00±0.00**	44.00±0.00

The best model

Best ESUNN model

$f_1(x, \theta_1)=4.69-18.69*SU_1$

$f_2(x, \theta_2)=9.3-13.62*SU_2$

$f_3(x, \theta_3)=0.49+13.49*SU1-23.59*SU2$

$f_4(x,\theta_4)=0$

$$g_l(\mathbf{x},\boldsymbol{\theta}_l)=\frac{\exp\ f_l(\mathbf{x},\boldsymbol{\theta}_l)}{\sum_{l=1}^{Q}\exp\ f_l(\mathbf{x},\boldsymbol{\theta}_l)} \text{ for } l=1,...,Q$$

$$SU_1=\frac{1}{(1+EXP-(-9.32+8.07*TERTIT+9.31*TRADEM-4.08*RLPGH+5.03*PATREV+3.68*HUMLF))}$$

$$SU_2=\frac{1}{(1+EXP-(-12.44+3.42*TERTIT-9.61*LLEARN-9.04*GERDGO-3.07*GDPGRO+7.32*TPATRE+4.47*HUMLF}$$

CCR_G= 100% ; MS_G= 100%; K_G=1

Variable	Probability Cluster 1 (a)	Probability Cluster 2	Probability Cluster 3	Probability Cluster 4
TERTIT	**(--)**	**(--)**	**(+)**	**(++)**
LLEARN	**(--)**	**(++)**	**(+)**	**(--)**
GERDGO	**(-)**	**(++)**	**(-)**	**(--)**
TRADEM	**(--)**	**(--)**	**(++)**	**(++)**
RLPGH	**(+)**	**(=)**	**(=)**	**(--)**
GDPGRO	**(=)**	**(=)**	**(=)**	**(=)**
TPATRE	**(--)**	**(++)**	**(+)**	**(--)**
PATREV	**(--)**	**(--)**	**(++)**	**(++)**
HUMLF	**(-)**	**(=)**	**(-)**	**(=)**

Conclusions

- **MANAGERIAL** TOOL: supporting decision making in R&D, since it provides information about strengths and weaknesses in innovation activities and about their contribution to the country's competitiveness
- **BENCHMARKING** TOOL: allowing the EU to compare its own innovation activities, inputs and results with other countries of reference and with European Members national initiatives
- **MONITORING** and **EVALUATION** TOOL for institutional and governmental bodies, since it allows some critical innovation indicators to be monitored.

EVOLUTIONARY NEURAL NETWORK CLASSIFIERS TO PREDICT THE CLASSIFICATION OF R&D PERFORMANCE IN EU COUNTRIES

THANK YOU FOR YOUR ATTENTION!

M. De la Paz-Marín, **P. Campoy-Muñoz**, C. Hervás-Martinez
ETEA, Business Administration Faculty, University of Cordoba
AYRNA Reserch Group, University of Córdoba
mpcampoy@etea.com

XVII SIGEF Congress "Methods for Decision Making in an uncertain environment"
6-8 Junio, Reus