



Clasificación de desastres y emergencias con representación bipolar del conocimiento

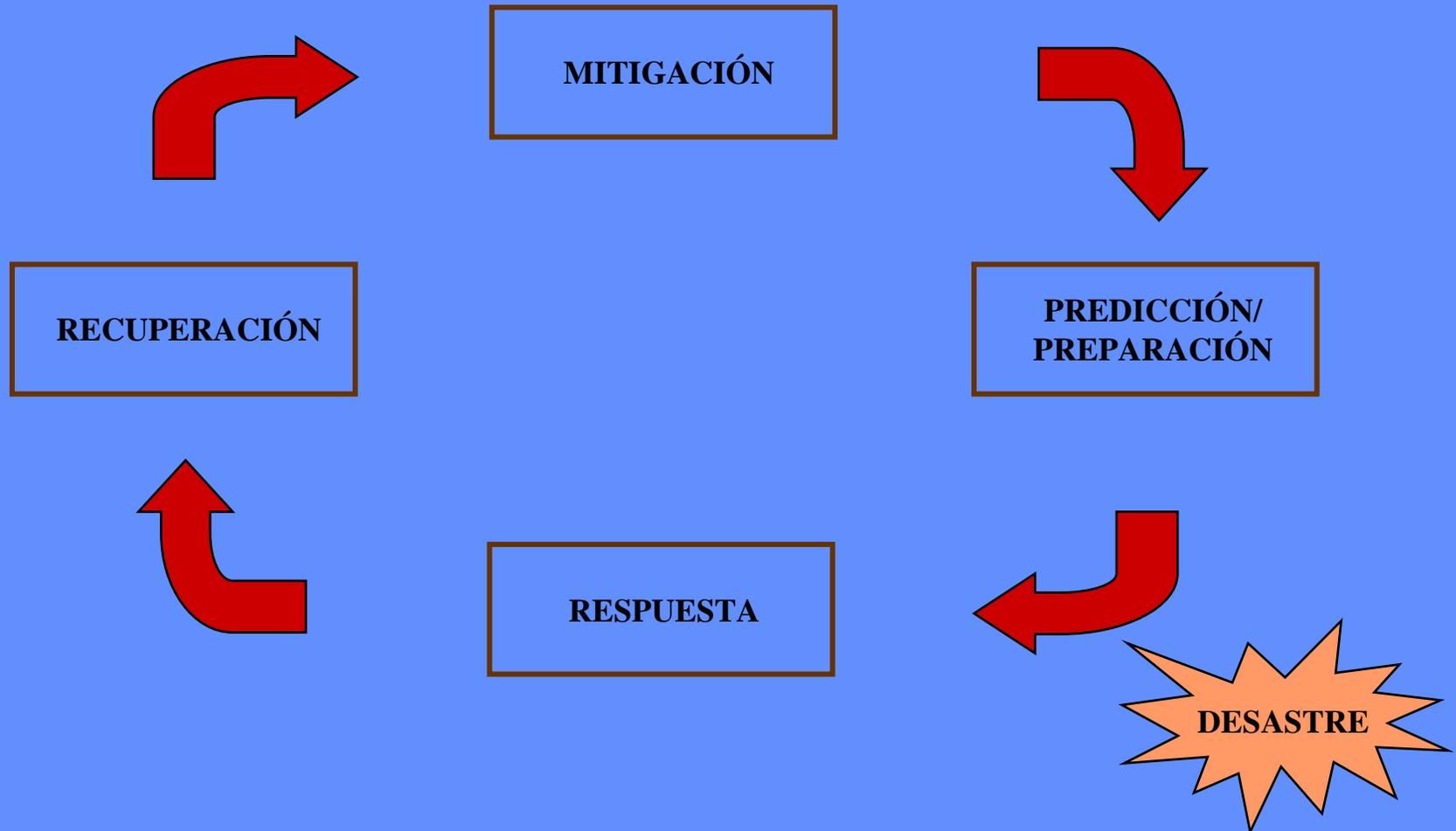
J. Tinguaro Rodríguez, Begoña Vitoriano, Javier Montero

Departamento de Estadística e Investigación Operativa
Universidad Complutense de Madrid

Madrid, 1 de febrero de 2011



FASES DE LA GESTIÓN DE DESASTRES





JERARQUÍA DE DECISIONES EN LAS ONG PARA LA RESPUESTA A DESASTRES





Problema general de evaluación de consecuencias de desastres:

- Distintas combinaciones de tipo de **desastre y localización**.
- Situación de **emergencia**.
- **Incertidumbre e imprecisión** de la información.
- **Restricciones** que impone el contexto de las ONG.

El **objetivo** de esta investigación es desarrollar un sistema de ayuda a la decisión con el que proporcionar ayuda a los decisores de las ONG en sus procesos de decisión estratégicos.

Este sistema de ayuda a la decisión ha sido denominado **Sistema Experto para el Diagnóstico de Desastres (SEDD)**.

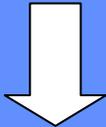
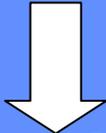
SEDD se centra en proponer una **evaluación** de las consecuencias de los desastres a partir de la **primera información disponible**.

Desastres, respuesta humanitaria e IA



- Para garantizar la mayor **aplicabilidad** de SEDD, la evaluación de un desastre bajo estudio será obtenida a partir de:
 - **información histórica** (conocimiento) y
 - variables descriptoras o **atributos** de ese desastre **fácilmente accesibles** (información).
- Información histórica: base de datos sobre desastres **EM-DAT** (Emergency Events Database, www.emdat.be):
 - **CRED** (Center for the Research on the Epidemiology of Disasters, www.cred.be) de la **Universidad de Lovaina** (Bélgica).
 - Registros de **más de 16.000 casos** desde el año 1900 hasta la actualidad.
 - Información sobre un **conjunto de consecuencias reducido pero significativo**



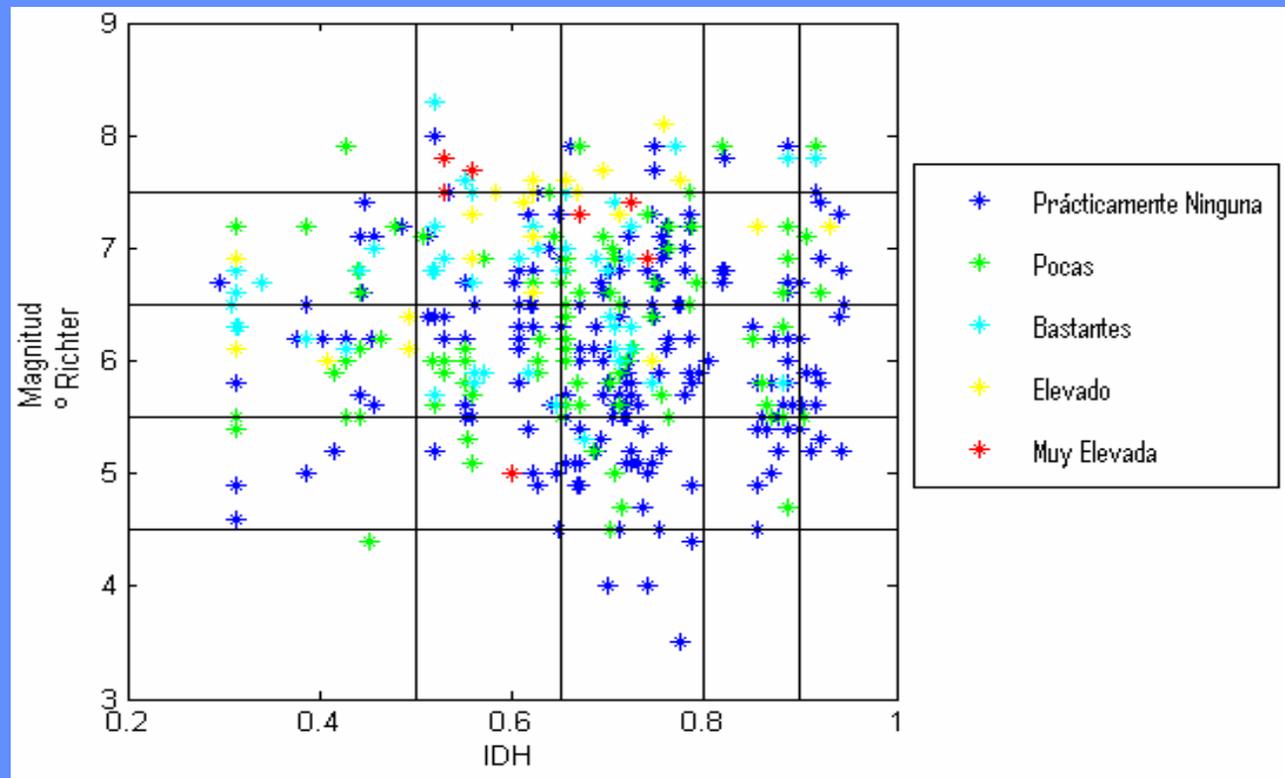
- **Valoración numérica precisa es poco realista:**
 - **Incertidumbre** de la información disponible en los primeros instantes tras la ocurrencia de un desastre
 - **Imprecisión** implícita de algunas de las variables en EM-DAT (ej. 50.000 afectados)
- Es más plausible **clasificar** las consecuencias de un desastre en términos de los **escenarios relevantes** para los decisores de las ONG.
- Problema de **clasificación supervisada:**
 - Las clases se identifican con los **escenarios relevantes**.
 - El conjunto de clases está dotado de **estructura (gradación)**.
 - Características específicas del contexto: Riesgo subestimación

Desastres, respuesta humanitaria e IA



- Problema de clasificación difícil porque EM-DAT presenta:
 - Gran **variabilidad**.
 - Fuerte **descompensación** entre clases.
 - **Carencia de variables explicativas** suficientes.

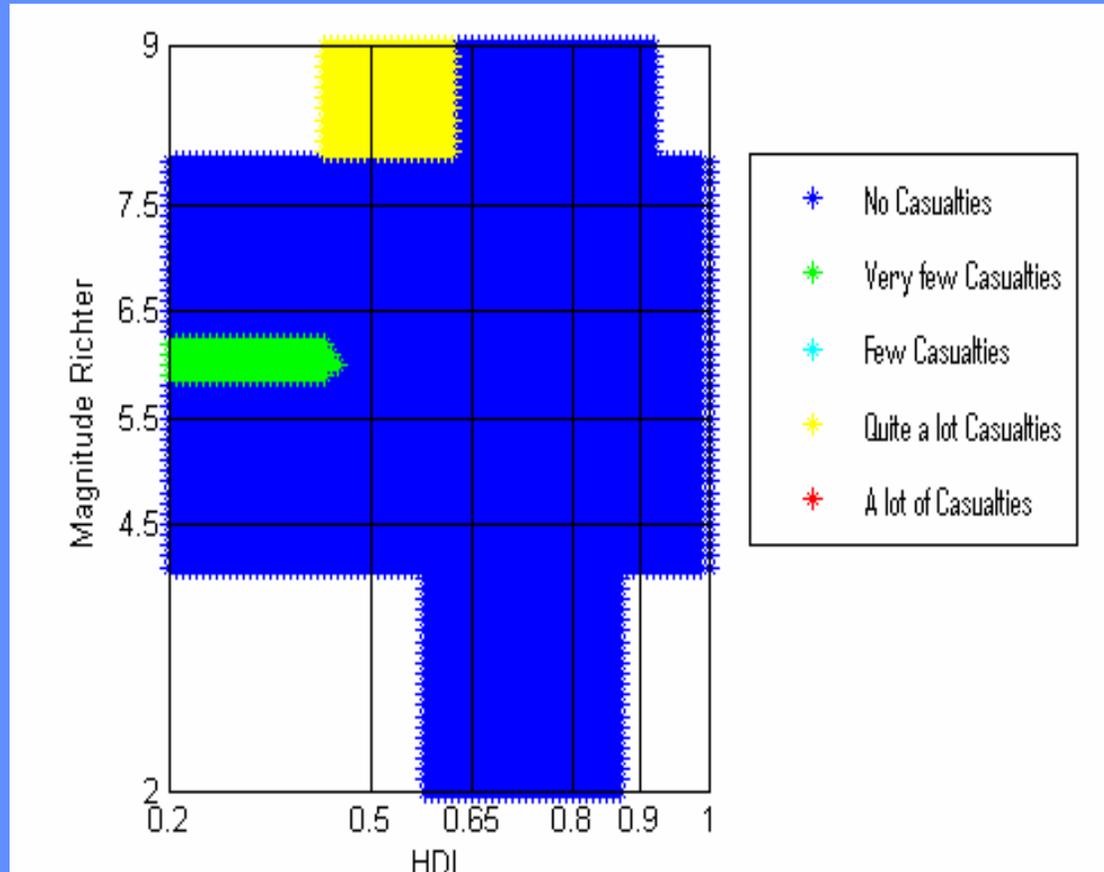
Figura 5.5: Muestra de entrenamiento para variable **Número víctimas mortales (NVM)** y **Tipo desastre = Terremoto**



- Complican la viabilidad estadística clásica (Rodríguez et al., 2010c).
Rodríguez, J.T., Vitoriano, B., Montero, J. (2010c) A disaster-severity assessment DSS comparative analysis, *OR Spectrum* (IN SECOND REVISION)

Desastres, respuesta humanitaria e IA

- Resultados obtenidos mediante un clasificador descriptivo estándar, como el propuesto en [Rodríguez et al., 2010b]:



Rodríguez, J. T., Vitoriano, B., Montero, J. (2010b) A general methodology for data-based rule building and its application to natural disaster management, *Computers & Operations Research*, In Press, Corrected Proof



- Falta de adaptación de estos resultados.
- Características no consideradas:
 - **Estructura** del conjunto de clases.
 - Asimetría derivada del riesgo de **subestimación**.
- Estas consideraciones conducen a la noción de ***bipolaridad*** → introducción de **información negativa**.
- La introducción de un enfoque bipolar plantea:
 - Necesidad de definición precisa de la **información negativa**.
 - **Modelos lógicos** de representación del conocimiento más complejos.
 - Procedimientos de **aprendizaje** que permitan la utilización de información negativa.
 - Métodos de **razonamiento** que admitan grados de confianza bipolares.



Clasificación bipolar

Ejemplo:

R: si un terremoto de intensidad *moderada* ocurren en un país con *IDH medio-alto* entonces *NVM = prácticamente ninguna*.

Una instancia que cumpla esa premisa y verifique *NVM=muy elevado* es tan contraejemplo como otra con *NVM=pocos*.

- Distinción de los contraejemplos entre **excepciones menores** y **excepciones significativas** → menor confianza de las reglas.
- Esto se lleva a cabo introduciendo una **matriz de disimilaridad** $\Delta = (d_{ij})_{N_C \times N_C}$ en el conjunto de clases o **consecuentes** de manera que d_{ij} representa el grado en que la clase C_j es disimilar a la clase C_i , por lo que $d_{ij} \in [0,1], d_{ii} = 0 \forall i$ y

$$\mu_{\partial C_j}(y) = \sum_{i=1, \dots, N_C} d_{ji} \cdot \mu_{C_j}(y)$$

representa el grado en que y constituye información negativa para la clase C_j .



Clasificación bipolar

- Se propone evaluar las **evidencias positiva** r_j^+ (a favor) y **negativa** r_j^- (en contra), de una regla $R_j(A) : A \Rightarrow C_j$, mediante las siguientes expresiones:

$$r_j^+(A) = \frac{\sum_{p=1}^m \mu_A(x^p) \cdot \mu_{C_j}(y^p)}{\sum_{p=1}^m \mu_A(x^p)} = cf(A \Rightarrow C_j)$$

$$r_j^-(A) = \frac{\sum_{p=1}^m \mu_A(x^p) \cdot \mu_{\partial C_j}(y^p)}{\sum_{p=1}^m \mu_A(x^p)} = cf(A \Rightarrow \partial C_j)$$

- Nótese que $r^+, r^- \in [0,1]$ y que $r^+ + r^- \leq 1$.
- Este grado de confianza bidimensional puede ser agregado en un valor unidimensional de diversas maneras, por ejemplo

$$t_j(A) = \max\{r_j^+(A) - r_j^-(A), 0\}$$



Clasificación bipolar de desastres

- La aplicación del modelo de clasificación bipolar recién propuesto al problema de SEDD supone la **búsqueda de matrices de disimilaridad** que permitan considerar e introducir los **requisitos** del contexto y del problema en el modelo de clasificación, en particular en el **aprendizaje** de reglas de clasificación a partir de los datos de EM-DAT.
- **Medidas de error** alternativas que complementen a las tradicionales medidas de eficacia predictiva (%CC, %NC, %CT, %ERROR) y completen la **descripción cuantitativa** proporcionada por éstas.



Clasificación bipolar de desastres

Medidas de error complementarias

- Media $\overline{PRED} = \frac{\sum_{p=1...m} C_{PRED}^p}{m_C}$
- Varianza $VAR_{PRED} = \frac{\sum_{p/C_{PRED}^p \neq 0} (C_{PRED}^p - \overline{PRED})^2}{m_C}$
- Diferencia de proporciones $DIFPROP = \sum_{j=1...N_C} |PROP_j^R - PROP_j^P|$
- Coste $COST = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 2 & 0 & 1 & 2 & 3 \\ 4 & 2 & 0 & 1 & 2 \\ 6 & 4 & 2 & 0 & 1 \\ 8 & 6 & 4 & 2 & 0 \end{bmatrix} \rightarrow COSTE = \frac{\sum_{p/C_{PRED}^p \neq 0} COST_{C_{REAL}^p, C_{PRED}^p}}{m_C}$
- Además, el comportamiento de cada clasificador es **simulado** sobre un retículo denso del espacio de entrada, lo que permitirá obtener descripciones **cualitativas** de esos comportamientos.

Clasificación bipolar de desastres

Resultados de los clasificadores no bipolares

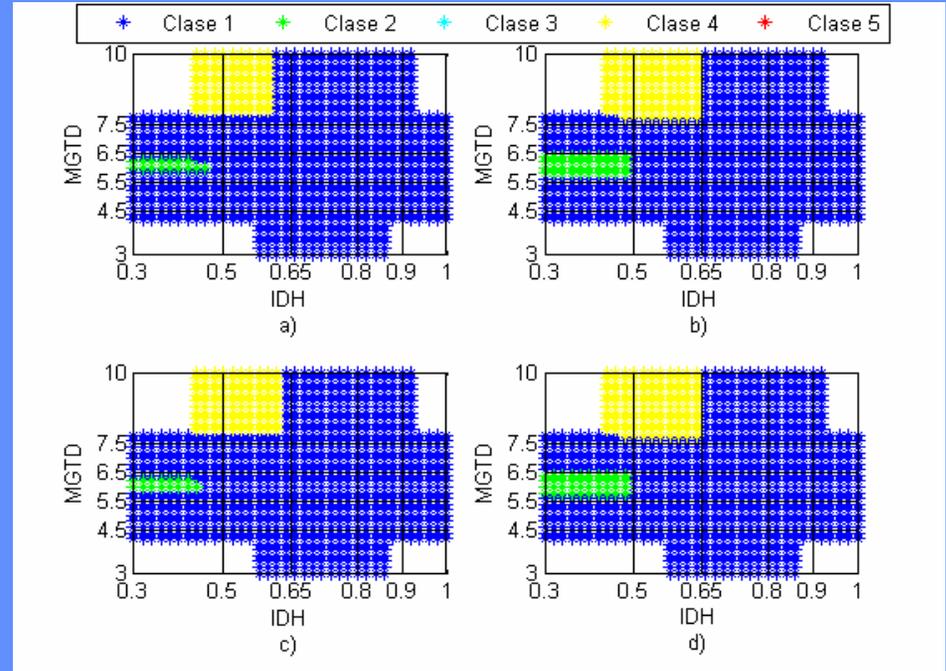
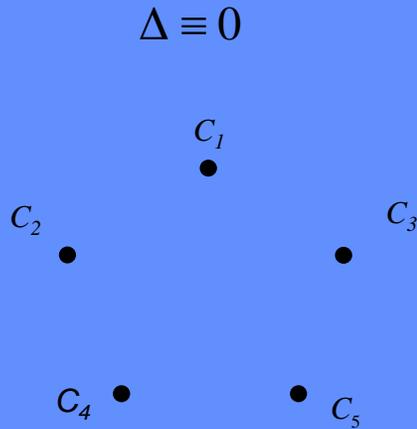


Figura 5.7 y Tablas 5.1 y 5.2

Método	%CC	%NC	%CT	%ERROR	COSTE	PRED	VAR	DIFPROP
Validación completa								
Nítido+AN	54,15	0	54,15	45,85	1,4534	1,0415	0,0708	0,8497
Nítido+RG	53,37	0	53,37	46,63	1,4119	1,0777	0,1339	0,7979
Híbrido+AN	54,15	0	54,15	45,85	1,4534	1,0415	0,0708	0,8497
Híbrido+RG	53,37	0	53,37	46,63	1,4119	1,0777	0,1339	0,7979
Validación cruzada								
Nítido + AN	52,33	0	52,33	47,67	1,4819	1,0466	0,0807	0,8446
Nítido + RG	50	0	50	50	1,4715	1,1295	0,2527	0,7565
Híbrido + AN	52,85	0	52,85	47,15	1,4741	1,0389	0,0684	0,8549
Híbrido + RG	52,59	0	52,59	47,41	1,4352	1,0777	0,1494	0,8083



Clasificación bipolar de desastres

- La adaptación de estos métodos al contexto no es buena:
 - Grave **riesgo de subestimación**
 - **Transiciones bruscas**
 - **Pésima replicación de la muestra original**

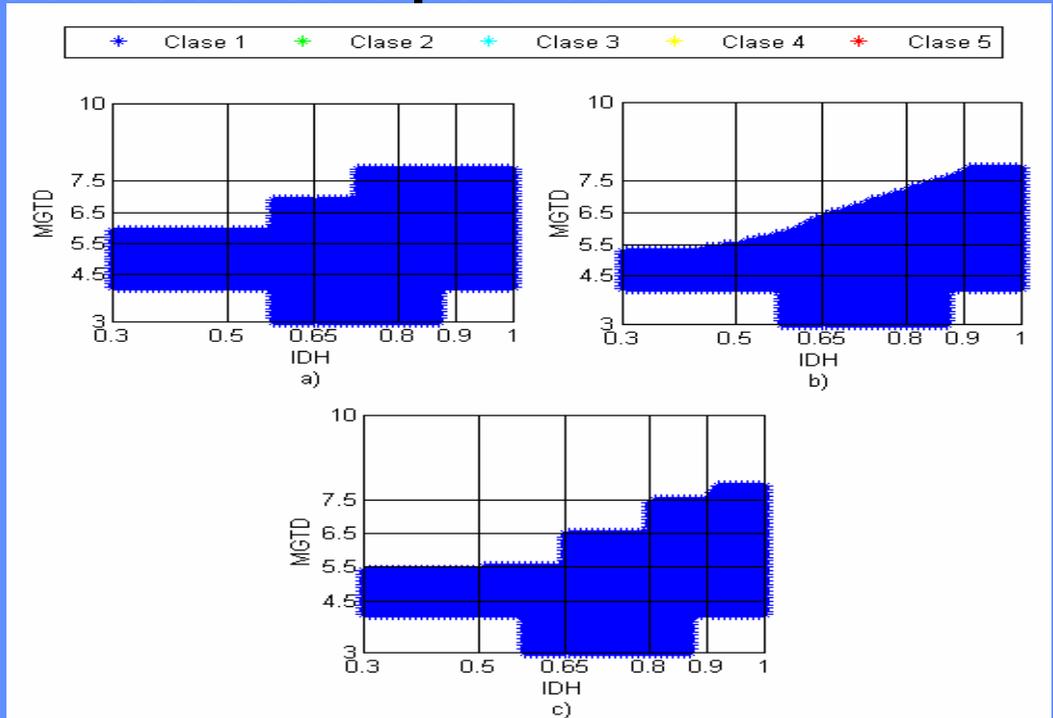
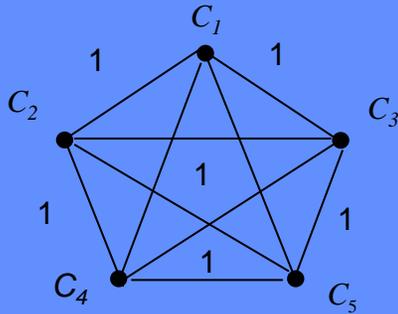
Pertencientes a	Clasificados en						Total	%
	<i>PN</i>	<i>P</i>	<i>B</i>	<i>E</i>	<i>ME</i>	<i>NC</i>		
<i>Prácticamente Ninguna (PN)</i>	206	3	0	1	0	0	210	54,4
<i>Pocas (P)</i>	97	3	0	0	0	0	100	25,91
<i>Bastantes (B)</i>	45	2	0	1	0	0	48	12,44
<i>Elevado (E)</i>	20	2	0	0	0	0	22	5,7
<i>Muy Elevado (ME)</i>	6	0	0	0	0	0	6	1,55
Total	374	10	0	2	0	0	386	
%	96,89	2,59	0	0,52	0	0		

Tabla 5.3

Clasificación bipolar de desastres

Resultados de los clasificadores bipolares

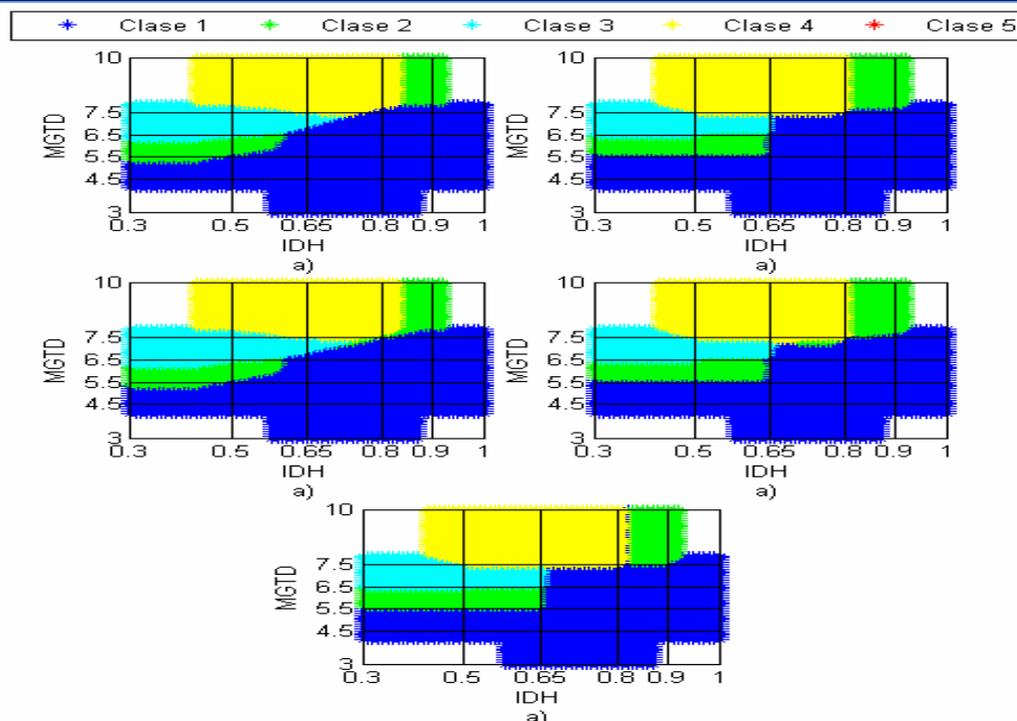
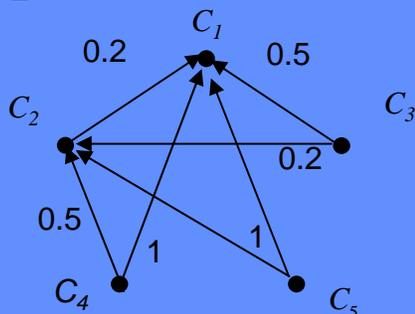
$$\Delta = \Delta^{K_5} = 1 - Id$$



Método	%CC	%NC	%CT	%ERROR	COSTE	PRED	VAR	DIFPROP
Validación completa								
VA1	63,8	27,72	46,11	26,17	0,9821	1	0	0,9119
VA2	69,12	43,78	38,86	17,36	0,8111	1	0	0,9119
MP2	71,13	49,74	35,75	14,51	0,7526	1	0	0,9119
Validación cruzada								
VA1	63,8	27,72	46,11	26,17	0,9821	1	0	0,9119
VA2	67,58	43,26	38,34	18,4	0,8584	1	0	0,9119
MP2	71,13	49,74	35,75	14,51	0,7526	1	0	0,9119

Clasificación bipolar de desastres

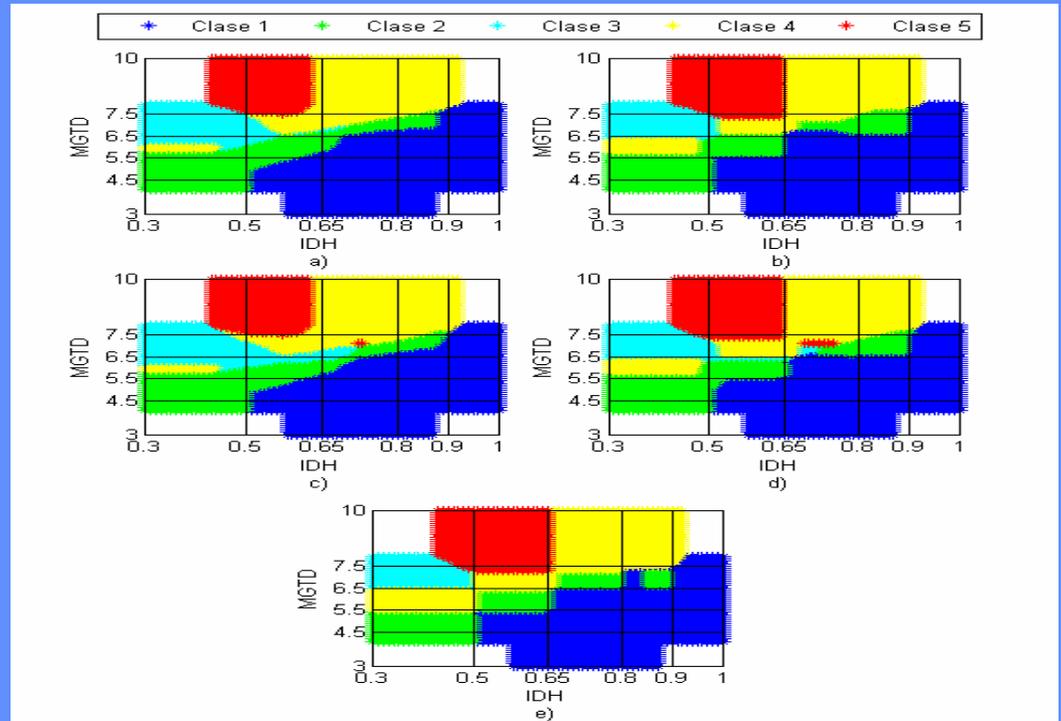
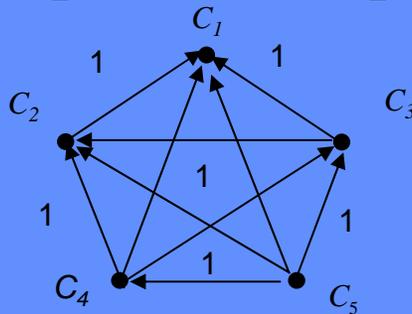
$$\Delta_3^{SUB} = \begin{bmatrix} 0 & 0.2 & 0.5 & 1 & 1 \\ 0 & 0 & 0.2 & 0.5 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Método	%CC	%NC	%CT	%ERROR	COSTE	PRED	VAR	DIFPROP
Validación completa								
VA1	55,18	0	55,18	44,82	1,0699	1,5518	0,7965	0,4093
MV1	53,37	0	53,37	46,63	1,0518	1,5881	0,8795	0,2591
VA2	55,18	0	55,18	44,82	1,0363	1,5803	0,8239	0,3938
MV2	52,85	0	52,85	47,15	1,0415	1,6244	0,9081	0,228
MP2	53,11	0	53,11	46,89	1,0415	1,6321	0,9268	0,2539
Validación cruzada								
VA1	51,55	0	51,55	48,45	1,1321	1,5518	0,7965	0,3886
MV1	51,81	0	51,81	48,19	1,0959	1,5699	0,8461	0,2694
VA2	51,04	0	51,04	48,96	1,1192	1,5699	0,8047	0,3731
MV2	51,04	0	51,04	48,96	1,0803	1,6166	0,8944	0,2435
MP2	51,04	0	51,04	48,96	1,0829	1,6347	0,9158	0,2591

Clasificación bipolar de desastres

$$\Delta_{SUB}^{ODL} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Método	%CC	%NC	%CT	%ERROR	COSTE	PRED	VAR	DIFPROP
Validación completa								
VA1	50,52	0	50,52	49,48	0,9689	2,0181	1,339	0,2176
MV1	48,19	0	48,19	51,81	1,0518	2,0466	1,6299	0,2953
VA2	47,93	0	47,93	52,07	1,0104	2,1295	1,4081	0,3316
MV2	46,63	0	46,63	53,37	1,0855	2,1813	1,7599	0,3575
MP2	45,34	0	45,34	54,66	1,0907	2,2254	1,8067	0,4093
Validación cruzada								
VA1	46,89	0	46,89	53,11	1,0259	2,0285	1,3748	0,2176
MV1	44,56	0	44,56	55,44	1,1114	2,0285	1,6028	0,2798
VA2	43,78	0	43,78	56,22	1,0751	2,1399	1,4416	0,3264
MV2	43,52	0	43,52	56,48	1,1451	2,171	1,7376	0,3472
MP2	40,67	0	40,67	59,33	1,1788	2,2047	1,769	0,3782



Clasificación bipolar de desastres

En conclusión, la aplicación de un enfoque bipolar :

- Permite caracterizar mejor la problemática planteada
 - **estructura de orden** del espacio de valuación del problema;
 - riesgo de **subestimación**.
- Produce clasificadores que mejoran consistentemente el rendimiento de los métodos no bipolares
 - en términos de **eficacia predictiva**;
 - confiables desde el punto de vista de la **probabilidad global de predecir cada clase**;
 - confiables desde el punto de vista de la **estabilidad** y la **continuidad** de sus predicciones.
- Existe una diferencia **cualitativa** entre el comportamiento de los modelos no bipolares y los clasificadores bipolares.
- Propician la emergencia de **conocimiento relevante** y no obvio en condiciones difícilmente tratables por los modelos estándar.



Gracias por su atención!!