

PRESENTACIÓN E IMPLEMENTACIÓN DEL MÉTODO DE NUBES DINÁMICAS DIFUSO

WALTER MORA FLORES¹

Resumen

Se hace una presentación del método de Nubes Dinámicas Difuso considerando la variante del centro de gravedad (Forgy, 1965). La clasificación difusa obtenida se usa como herramienta de validación para establecer el número k de clases.

Palabras clave: conjuntos difusos, clasificación por particiones, clasificación difusa, Forgy, validación de una clasificación.

1. Introducción

Entre los numerosos métodos que se han propuesto para clasificar un conjunto de objetos, están los métodos difusos. En este artículo se presenta el método de clasificación conocido como Nubes Dinámicas difuso, que permite a cada objeto pertenecer a diferentes clases simultáneamente según una función de pertenencia análoga a la que se usa para los conjuntos difusos. Se ilustra el método detalladamente a través de un ejemplo y se presentan los resultados obtenidos sobre varias tablas de datos.

Sean X, A conjuntos, $A \subseteq X$ y $\mu_A : X \rightarrow [0, 1]$. Se define el *conjunto difuso* \tilde{A} sobre X como

$$\tilde{A} = \{(a, \mu_A(a)) \mid a \in X\}.$$

μ_A se le llama *función de pertenencia* de \tilde{A} . En general, \tilde{A} es especificado en lenguaje natural como un concepto difuso; $\mu_A(a)$ mide el porcentaje en que los elementos del conjunto soporte X , cumplen este concepto. Si \tilde{A} no es un concepto difuso, se dice que \tilde{A} es clásico; en este caso $\mu_A = \chi_A$ la función característica de A .

Una exposición general sobre conjuntos difusos y sus aplicaciones se puede encontrar en [?] o en [?].

¹DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA

2. Clasificación clásica y clasificación difusa

Sea $\Omega = \{x_1, x_2, \dots, x_n\}$, el conjunto de objetos, ‘individuos’ o ‘patrones’ a clasificar. La clasificación por particiones es un proceso no supervisado de agrupamiento en k clases de los objetos de Ω . Formalmente, el proceso de clasificación en el sentido clásico, puede ser establecido así: encontrar k clases C_1, C_2, \dots, C_k , ($2 \leq k < n$) tales que:

1. $\Omega = \bigcup_{i=1}^k C_i$
2. $C_l \cap C_h = \emptyset$ para todo $l \neq h$

Los algoritmos de clasificación agrupan objetos en cada clase vía una asociación ‘natural’, de acuerdo con algún criterio que mide la calidad de la partición, usualmente basada en una medida de similitud o disimilitud calculada sobre los individuos (por ejemplo, Fory, k-means, nubes dinámicas, etc) [?]. Se espera que el grado de asociación sea alto entre elementos de la misma clase y bajo entre miembros de clases distintas (i.e. las clases deben ser homogéneas).

Uno de los criterios usados para clasificar es optimizar una función objetivo. Muchas funciones objetivo han sido sugeridas, tanto para clasificación clásica como para clasificación difusa [?]. El algoritmo que vamos a describir se conoce como *Nubes Dinámicas* o de Fory [?] en sus versiones clásica y difusa [?].

2.1. Método de Nubes Dinámicas, variante del centro de gravedad

La medida de disimilitud que usaremos es la distancia euclídea clásica:

$$d(x_i, x_j) = \|x_i - x_j\| = \left(\sum_{t=1}^p (x_i^t - x_j^t)^2 \right)^{\frac{1}{2}}$$

donde $x_i = (x_i^1, \dots, x_i^p) \in \Omega$.

Cualquier partición de Ω en k clases puede ser descrita por la familia de funciones características χ_{C_l} vía la matriz $U = [u_{li}]$ donde $u_{li} := \chi_{C_l}(x_i)$.

Ejemplo 1. Sea $\Omega = \{x_1, x_2, x_3, x_4\}$, $k = 2$, $C_1 = \{x_1, x_2, x_3\}$ $C_2 = \{x_4\}$. Entonces

$$U = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline C_1 & 1 & 1 & 1 & 0 \\ \hline C_2 & 0 & 0 & 0 & 1 \end{array}$$

La matriz U cumple las siguientes 3 condiciones:

1. Para todo $1 \leq l \leq k$, $1 \leq i \leq n$, $u_{li} \in \{0, 1\}$
2. Para todo $1 \leq i \leq n$, $\sum_{l=1}^k u_{li} = 1$

3. Para todo $1 \leq l \leq k$, $0 < \sum_{i=1}^n u_{li} < n$

La condición 2 indica que cada x_i está en una sola clase y la condición 3 significa que cada clase C_l tiene al menos un individuo. Durante el proceso de creación de una partición óptima, ninguna clase puede quedar vacía.

Sea $G = (g_1, g_2, \dots, g_k)$, $g_l \in \mathbb{R}^p$, la función objetivo es una inercia:

$$W(C_1, C_2, \dots, C_k, G) = \sum_{l=1}^k \sum_{i=1}^n u_{li} \|x_i - g_l\|^2$$

Por el teorema de Huygens ([?]), dada U , los centros de gravedad $g_l = \frac{1}{\text{card}(C_l)} \sum_{x_j \in C_l} x_j$ son los valores apropiados de G para minimizar esta suma, por lo que una clase se puede representar por su centro de gravedad. Así el criterio de clasificación se reduce a encontrar una matriz U^* que minimice W i.e. encontrar una partición que minimice la suma de las inercias de todas las clases; formalmente:

CRITERIO DE CLASIFICACIÓN: Minimizar $W(U) = \sum_{l=1}^k \sum_{i=1}^n u_{li} \|x_i - g_l\|^2$.

El anterior criterio plantea un problema de optimización combinatoria. El método de nubes dinámicas, en la variante del centro de gravedad (originalmente propuesto por Forgy en 1965) es convergente y aproxima el W mínimo en un número finito de pasos. El algoritmo, en breve, procede así:

ALGORITMO FORGY CLÁSICO

0. Escoja k , una distancia, un umbral ϵ de parada y k elementos de Ω que representen una partición inicial, estos se tomarán como los centros de gravedad iniciales.
1. Asigne cada x_i a la clase representada por el centro de gravedad más cercano.
2. Calcule W , $W_1 = W$.
3. Calcule los centros de gravedad de las clases formadas.
4. Calcule W .
5. Deténgase si $\frac{|W_1 - W|}{W} < \epsilon$, sino, vaya al paso 1.

El algoritmo anterior se llama a veces erróneamente k -means [?]. En la formulación de McQuenn de este último, en 1967, se calcula el centro de gravedad después de cada asignación, lo que introduce un sesgo dado por el orden en que se presentan los datos; esto corresponde al método de transferencias de S. Régnier.

2.2. Método de Nubes Dinámicas Difuso

Los algoritmos de clasificación clásicos generan particiones en las que cada objeto de Ω es asignado a una única clase. Sin embargo, a veces los objetos no pueden ser asignados estrictamente a una clase pues están localizados ‘entre’ distintas clases. En estos casos, los métodos de clasificación difusa representan, de una manera más adecuada, la situación real de los datos. En vez de asignar un objeto a una sola clase, lo que se hace es asignar el objeto a varias las clases en cierto porcentaje.

Definición: Una k -partición difusa de Ω es una matriz $\tilde{U} = (\mu_{li})_{k \times n}$ que satisface las siguientes condiciones:

1. Para todo $1 \leq l \leq k$, $1 \leq i \leq n$, $\mu_{li} \in \{0, 1\}$

2. Para todo $1 \leq i \leq n$, $\sum_{l=1}^k \mu_{li} = 1$

3. Para todo $1 \leq l \leq k$, $0 < \sum_{i=1}^n \mu_{li} < n$

La condición 2 indica que el grado de pertenencia de cada individuo está distribuido porcentualmente y en su totalidad, en las k clases; y la condición 3 significa que cada clase tiene un porcentaje estrictamente positivo de algún individuo y que un individuo no puede estar asignado a más clases que las que existen de antemano.

Ejemplo 2. Sea $\Omega = \{x_1, x_2, x_3, x_4\}$ $k = 2$, una posible 2-partición difusa de Ω puede ser

$$\tilde{U} = \begin{pmatrix} 0,8 & 0,999 & 0,9 & 0,01 \\ 0,2 & 0,001 & 0,1 & 0,99 \end{pmatrix} = \begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline \tilde{C}_1 & 0.8 & 0.999 & 0.9 & 0.01 \\ \tilde{C}_2 & 0.2 & 0.001 & 0.1 & 0.99 \end{array}$$

El criterio que se usará para clasificar es llamado *inercia difusa*; se trata, generalizando la inercia, de minimizar una función objetivo. Hay que encontrar una matriz \tilde{U}^* que resuelva el problema:

$$\text{Minimizar } W(\tilde{U}, \tilde{G}) = \sum_{l=1}^k \sum_{i=1}^n (\mu_{li})^m \|x_i - \tilde{g}_l\|^2, \quad m > 1$$

donde $\tilde{G} = (\tilde{g}_1, \dots, \tilde{g}_k)$, con $\tilde{g}_l \in \mathbb{R}^p$.

Aquí los x_i con alto grado de pertenencia en C_l , tienen una gran influencia sobre \tilde{g}_l . Esto es fortalecido con m . La función de m es reducir la influencia de los x_i ‘muy’ alejados de \tilde{g}_l en el cálculo de los \tilde{g}_l .

Derivando W respecto a \tilde{g}_l (para \tilde{U} fija) y aplicando la condición $\sum_{i=1}^n \mu_{li} = 1$ se obtiene:

$$\tilde{g}_l = \frac{1}{\sum_{i=1}^n (\mu_{li})^m} \sum_{i=1}^n (\mu_{li})^m x_i \quad (1)$$

como condición necesaria para que W sea mínima. Por otra parte, derivando W respecto a μ_{li} (para \tilde{G} fijo) se obtiene:

$$\mu_{li} = \frac{\left(\frac{1}{\|x_i - \tilde{g}_l\|^2}\right)^{1/(m-1)}}{\sum_{h=1}^k \left(\frac{1}{\|x_j - \tilde{g}_h\|^2}\right)^{1/(m-1)}} \quad (2)$$

μ_{li} resulta ser una ponderación de la atracción que ejerce \tilde{g}_l sobre x_j respecto a las atracciones de los otros \tilde{g}_h . Si x_i está cerca de \tilde{g}_l pero suficientemente lejos de los otros \tilde{g}_h , entonces² $\mu_{li} \approx 1$.

Las condiciones (1) y (2) son condiciones necesarias para alcanzar un óptimo local. El sistema descrito por (1) y (2) no se puede resolver analíticamente pero soluciones aproximadas pueden ser obtenidas usando un algoritmo muy similar al de nubes dinámicas. El siguiente algoritmo fue propuesto por Bezdek [?].

ALGORITMO FORGY DIFUSO

0. Escoja k y m , una distancia, una norma matricial Δ y un umbral $\epsilon > 0$. Inicialice \tilde{U}_0 ; $iter := 0$.
1. Calcule $\tilde{G} = (g_1, \dots, \tilde{g}_k)$ usando la ecuación (1)
2. Calcule \tilde{U}_{iter+1} usando la ecuación (2). En el caso de que algún $x_i = \tilde{g}_l$; defina $\mu_{ji} = 1$ si $j = i$ y $\mu_{ji} = 0$ si $j \neq i$.
3. Calcule $\Delta = \|\tilde{U}_{iter+1} - \tilde{U}_{iter}\|$.
Si $\Delta < \epsilon$ pare, sino $iter := iter + 1$ y vuelva al paso 1.

Usualmente se usa $m = 2$ y \tilde{U}_0 se inicializa con el resultado U , de una ejecución de ForgY clásico. Los valores grandes de m tienden a bajar los valores de μ_{li} (de hecho $1 \geq \max \mu_{li} \geq 1/k$). Las matrices \tilde{U} preferidas son las que tienen grados altos de pertenencia asignados a una clase pues esto indica alta concentración de los puntos alrededor de su centro de gravedad, lo que evidencia que las clases están razonablemente separadas. Si se escoge un k apropiado y si existe una buena estructura en los datos para clasificar, las particiones generadas por ForgY difuso son estables [?].

3. Validación

Para medir la validez en clasificación difusa, se han adaptado criterios tomados de clasificación clásica. También están los llamados *funcionales de validación* los cuales ayudan a escoger el número k .

Algunos criterios heurísticos que se usan son [?]:

²Para ver esto, observe que el numerador está presente en la suma del denominador

- a) *Método del codo*: el valor de la función objetivo W decrece si k aumenta, alcanza su mínimo cuando $k = n$. Si la función objetivo tuvo un ‘gran’ decrecimiento cuando se pasó de k^* a $k^* + 1$, se escoge $k = k^* + 1$.
- b) *Independencia’ de la inicialización*: la ‘estabilidad’ de la partición final respecto a diferentes inicializaciones puede ser usado como indicador de un k ‘correcto’.
- c) *Funcionales de validación*: estos funcionales dan el grado de ‘separabilidad’ entre clases (o el grado de difusidad de la partición). Se hacen clasificaciones para $k=2,3,\dots$; el número ‘correcto’ k es aquel asociado al funcional que da el mejor valor. En [?] se comparan 8 funcionales propuestos por diversos autores. Se señala como el de mejor desempeño al funcional denotado con $NF''I$:

$$NF''I = \frac{kF'' - 1}{k - 1}$$

$$\text{donde } F'' = \frac{1}{2} \left(\sum_{i=1}^N \frac{\mu_{x_i}}{N} + \min_{x_i} \{ \mu_{x_i} \} \right), \text{ aquí } \mu_{x_i} = \max\{ \mu_{1i}, \mu_{2i} \dots \mu_{ki} \}.$$

Si las clases C_1, C_2, \dots, C_k están ‘muy separadas’ entonces $\mu_{x_i} \approx 1$, con lo que $F'' \approx 1$ y entonces $NF''I \approx 1$. El peor caso es si $\max \mu_{li} \approx \frac{1}{k}$; en este caso se tendría $F'' \approx \frac{1}{k}$ con lo que $NF''I \approx 0$. El $\min_{x_i} \mu_{x_i}$ evidencia el grado de traslape (o separación) entre clases. En [?] se muestra cómo para diferentes valores de m , $NF''I$ detecta el mejor k (conocido a priori) en varios experimentos de clasificación con datos a los que se les aplicó Fuzzy difuso.

4. Resultados

4.1. Representación de la matriz \tilde{U}

En datos reales no siempre es posible contruir una barrera entre dos clases que se traslapan. En estos casos es de mucha ayuda visualizar la información disponible sobre estos traslapes. Una manera de lograr esta visualización es hacer una ‘proyección’ de la matriz \tilde{U} sobre el plano, que ponga en evidencia los datos bien clasificados y los datos que están en la zona de los traslapes, para cada valor de k .

Para hacer una representación de $\tilde{U}_{k \times n} = (\mu_{li})$ consideramos a la columna j como una representación de x_j . De esta columna nos interesa particularmente el máximo valor de μ_{lj} ($l = 1, \dots, k$), denotado por Φ_j . El resto de la columna puede ser vista como el ‘traslape’ de x_j . La representación de x_j será el vector $z_j = \left(\Phi_j, \sum_{i=1}^k \mu_{i,j} - \Phi_j \right) = (\Phi_j, \Upsilon_j)$. Estos puntos están sobre la recta $y = 1 - x$ por lo que, para efectos de visualización, se plotea el par $(\Phi_j, \Upsilon_j + \text{Random}[\pm] * \Upsilon_j * \text{Random}[0, 0.5])$. Esto no afecta la visualización puesto que si x_j está ‘bien clasificado’, entonces Υ_j es muy pequeño; en el otro caso, acentúa el traslape de x_j .

En los siguientes ejemplos con datos reales y simulados, para visualizar la información disponible, se comparan el método gráfico (representación de \tilde{U}), el índice $NF''I$ y el valor de la función objetivo W , esto con el propósito de establecer el ‘mejor’ k .

En todos los ejemplos se usó la distancia euclídea clásica con $m = 2$ y $\epsilon = 0,001$. Cada ejemplo se ejecutó varias veces sin que se notaran diferencias en el número k que sugería cada método.

4.2. Resultado con un conjunto de datos simulados.

Sea $\Omega = \{ (0.5, 1), (0.5, 2), (1, 1), (1, 2), (1, 3), (2, 2), (2.25, 3.4), (3, 2), (3.3, 3), (4, 2), (5, 1), (5, 2), (6, 1), (6, 3), (2, 5), (3, 4), (3, 5), (3, 6), (4, 4), (4, 5), (4, 6), (5, 6) \}$, representado en la figura 1.

Figura 1: Conjunto de datos simulados.

Los datos x_7, x_8, x_9 (representados con los puntos grandes) se han incluido para simular traslapes entre las 3 clases representadas con los otros puntos.

Ejecutando para valores de k de 1 hasta 5, se obtiene la información de la tabla de la figura 2.

k	NF' 'I	W
2	0.298965	50.9599
3	0.389656	22.206
4	0.459697*	14.4122
5	0.439858	10.7418

Figura 2: Tabla de resultados.

Figura 3: Método del codo.

La representación gráfica de \tilde{U} y la gráfica de W vs k , se dan en las figuras 4 y 3.

Como se observa, el método gráfico y el método del codo sugieren $k = 3$, mientras que el índice sugiere $k = 4$. En el método gráfico se evidencia la existencia de traslapes (los individuos esperados, por supuesto). La eliminación de estos tres individuos lleva

Figura 4: \tilde{U} para $k = 3$.Figura 5: \tilde{U} para $k = 4$.

a una coincidencia de los tres métodos en $k = 3$. En datos reales, la asignación de estos individuos requiere más análisis.

4.3. Iris de Fisher

Haciendo una ejecución para valores de k de 1 hasta 5, se obtienen los siguientes resultados:

k	NF''I	W
2	0.467666	117.685
3	0.624689*	51.4512
4	0.504152	36.8888
5	0.467352	37.8225

Figura 6: Tabla de resultados.

Figura 7: Método del codo.

En este caso, los tres métodos coinciden en $k = 3$. En la representación gráfica de \tilde{U} para $k = 3$, se observa la 'buena' clasificación que se logra para todos los individuos.

4.4. Peces de Amiard

Haciendo una ejecución para valores de k de 1 hasta 5, se obtiene la información que sigue:

En este caso, el índice $NF''I$ sugiere $k = 4$ por un estrecho margen sobre $k = 3$. La representación gráfica revela que $k = 3$ es escasamente mejor. El método del codo indica $k = 3$ sin ninguna duda. Los individuos que aparecen claramente separados de la mayoría de los puntos son $x_{11}, x_{12}, x_{13}, x_{14}$ y x_{24} , este patrón se mantiene en todas las ejecuciones que se hicieron, eventualmente aparecen un poco separados x_{18} y x_{20} . Si eliminamos de la

Figura 8: \tilde{U} para $k = 3$.

Figura 9: \tilde{U} para $k = 4$.

k	NF' 'I	W
2	0.362749	1173290
3	0.523252	550007
4	0.525797*	293121
5	0.444786	203207

Figura 10: Tabla de resultados.

Figura 11: Método del codo.

Figura 12: \tilde{U} para $k = 3$.

Figura 13: \tilde{U} para $k = 4$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}
,9	,9	,9	,9	,9	,9	,9	,9	,9	,9	,5	,2	,6	,7	,9	,9	,1	,0	,0	,8	,0	,0	,5
,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	,3	,1	,2	,1	,0	,0	,0	,9	,0	,0	,0	,9	,0
,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	,2	,6	,1	,1	,0	,06	,8	,0	,8	,0	,9	,0	,3

Cuadro 1: Matriz \tilde{U} para $k=3$ (redondeada).

tabla de datos a los individuos $x_{11}, x_{12}, x_{13}, x_{14}$ y x_{24} , los tres métodos coinciden en $k = 3$ de manera clara como se muestra en las figuras 14, 15 y 16.

k	NF' ' I	W
2	0.464719	814017
3	0.781041*	243704
4	0.566136	141874
5	0.479571	107533

Figura 14: Tabla de resultados.

Figura 15: Método del codo.

Figura 16: \tilde{U} para $k = 3$.

Figura 17: \tilde{U} para $k = 4$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{15}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
,01	,02	,05	,01	,06	,02	,03	,02	,01	,05	,04	,05	,06	,99	,00	,00	,93	,77
,98	,97	,90	,98	,90	,96	,96	,97	,99	,91	,95	,93	,91	,00	,00	,00	,04	,18
,00	,00	,03	,00	,03	,01	,01	,00	,00	,03	,01	,01	,02	,00	,98	,99	,02	,04

Figura 18: Matriz \tilde{U} para $k = 3$ (redondeada).

En la siguiente figura se muestra el plano principal del Análisis en Componentes Principales sobre la tabla de datos completa de los peces de Amiard, para las variables estandarizadas y la métrica identidad; los elementos de la clase 1 están representados con el signo \bullet mientras que los de las clases 2 y 3 con $*$ y $+$, respectivamente.

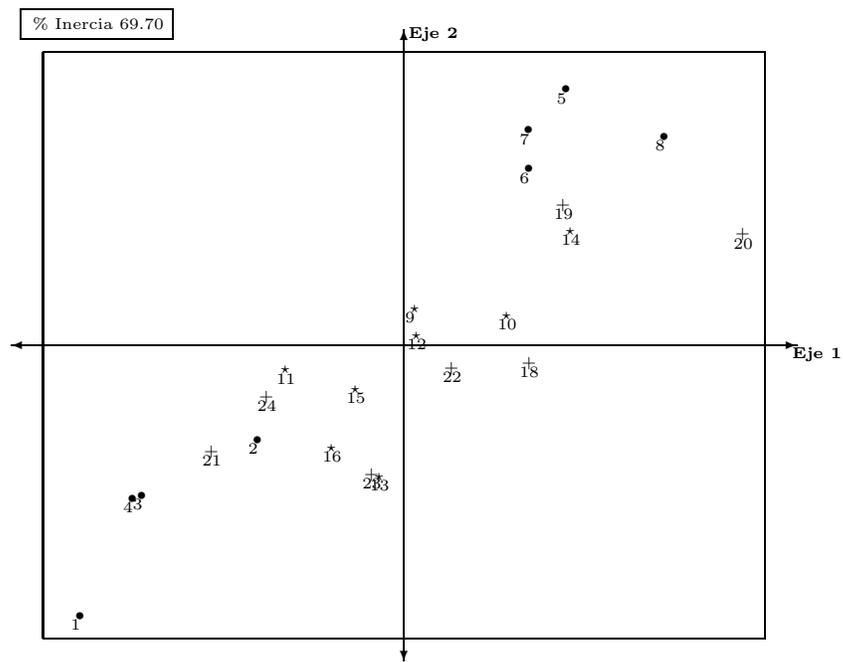


Figura 19: ACP para la tabla de datos de los peces de Amiard.

5. Conclusiones

En los ejemplos que se presentaron, y en otros experimentos, se vió claramente como \tilde{U} detecta traslapes entre las clases y como estos llegan a ‘confundir’ al índice NF^*I aunque sea escasamente. La representación de \tilde{U} mostró buen desempeño como herramienta de visualización de la clasificación en todas las corridas. La eliminación de los individuos que contribuyen al traslape hace que los tres métodos coincidan en el número k de manera clara. Estos individuos que se encuentran en los traslapes de las clases, posiblemente requieran un análisis individual para su clasificación.

Referencias

- [Zim85] Zimmermann, H.J. (1985) *Fuzzy Set Theory and its Applications*. Kluwer-Nijhoff Publishing, Boston.
- [Bez81] Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, London.
- [Did82] Diday, E. et al. (1982) *Eléments D’Analyse de Données*. Dunod, Paris.
- [Lib83] Libert, G.; Roubens, M. (1983) “New experimental results in cluster validity of fuzzy clustering algorithms,” *New Trends in Data Analysis and Applications*.
- [Asto94] Astorga, A.; Mora, W. (1994) “Clasificación de datos: implementación de un algoritmo clustering difuso”. *Tecnología en Marcha*, ITCR. Vol. 12.
- [Pao89] Pao, Y.-H. (1989) *Adaptive Pattern Recognition and Neural Networks*. Addison Wesley, Reading Mass..