



Inteligencia Artificial. Revista Iberoamericana
de Inteligencia Artificial

ISSN: 1137-3601

revista@aepia.org

Asociación Española para la Inteligencia
Artificial
España

Soto Espinosa, Jesús A.; Flores Sintas, Antonio; Vigo Aguiar, M. Isabel
Marco formal para una nueva función objetivo en agrupación difusa
Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 8, núm. 23, verano, 2004, p.
0
Asociación Española para la Inteligencia Artificial
Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=92502305>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Marco formal para una nueva función objetivo en agrupación difusa

Jesús A. Soto Espinosa*
Dpt. de Informática
jsoto@pdi.ucam.edu

Antonio Flores-Sintas*
Dpt. de Informática
aflores@pdi.ucam.edu

M. Isabel Vigo Aguiar**
Dpt. de Matemática Aplicada
vigo@ua.es

Resumen

Las técnicas de agrupación difusa trabajan, casi exclusivamente en la literatura existente, en la minimización de las distancias de los elementos de la muestra. La formulación de esta función objetivo se presenta exigiendo condiciones, como el grado difuso de la partición igual a 2, para que los resultados sean acordes con los esperados. En el presente trabajo damos marco formal a una deducción diferente de la función objetivo a partir de las propiedades locales de la muestra y cuyo significado físico parte de la *acción*, que conllevará una mejora en la interpretación de los grados de pertenencia y una concepción diferente de los algoritmos de agrupación difusa.

Palabras clave: Agrupación difusa, función objetivo, fuzzy clustering, fuzzy c-means.

1. Introducción

En general, las técnicas de agrupación (*clustering*) difusa se basan en encontrar el mínimo una función objetivo que determina los prototipos de los grupos buscados. El número de grupos suele ser un parámetro conocido c . Llamemos

$$v = \{v_1, v_2, \dots, v_c\}$$

a los prototipos buscados. Sea

$$X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^F$$

el conjunto de todos los elementos de la muestra, y llamemos u_{xk} a la probabilidad de que el elemento $x \in X$ pertenezca al grupo k . La función objetivo más utilizada en clustering, es la que nos mide las desviaciones respecto de los prototipos,

$$J = \sum_{k=1}^c \sum_{x \in X} u_{xk}^2 d_{xk}^2. \quad (1)$$

Fue Dunn [2] quien la introdujo al formalizar el algoritmo FCM. Aquí Dunn estudió dos funciones

$$J_1 = \sum_{k=1}^c \sum_{x \in X} u_{xk} \|x - v_k\|^2$$

y

$$J_2 = \sum_{k=1}^c \sum_{x \in X} u_{xk}^2 \|x - v_k\|^2.$$

La única diferencia entre las dos funciones objetivo es el exponente que afecta a las probabilidades de pertenencia, lo que lleva a Bezdek [1] a generalizar la función objetivo de la siguiente forma:

$$J_m = \sum_{k=1}^c \sum_{x \in X} u_{xk}^m \|x - v_k\|^2. \quad (2)$$

siendo m un exponente que regula el grado difuso "fuzzificación" de la partición. Si $m = 1$ la partición es 'hard', y para valores mayores la partición

*Universidad Católica San Antonio de Murcia.

**Universidad de Alicante.

es difusa; aumentando m la partición es más difusa.

Esta función será mínima cuando los prototipos estén situados de tal manera que representen agrupaciones de elementos alrededor de los prototipos. Si para $m=1$ Dunn encuentra que el mínimo sólo se puede existir si $u_{xk} = \{0, 1\}$, para $m = 2$, imponiendo la condición $\sum_{k=1}^c u_{xk} = 1$ encuentra que

$$v_k = \frac{\sum_{x \in X} u_{xk}^m x}{\sum_{x \in X} u_{xk}^m},$$

$$u_{xk} = \frac{\|x - v_k\|^{\frac{-2}{(m-1)}}}{\sum_{j=1}^c \|x - v_j\|^{\frac{-2}{(m-1)}}}, 1 \leq k \leq c$$

La función J_m está generalmente aceptada, aunque presenta una clara contradicción [5]: si aumentamos m , como las probabilidades son menores de la unidad, el valor de la función objetivo es cada vez menor, y esto nos lleva a que, para el mismo número de grupos, una partición más difusa tiene un error menor que una partición menos difusa.

Flores-Sintas [3] también refleja que no es solo el problema en relación con el exponente m , además el algoritmo FCM no funciona correctamente cuando los grupos no están demasiado alejados y tienen tamaños desproporcionados. Cuestiones que indujeron a Kamel y Selim [8], y, Krishnapuram [10, 9] a proponer varias alternativas empíricas para corregir estas deficiencias.

Otro punto a tener en cuenta sería la obtención de las probabilidades. Para la deducción de estas dadas por el FCM, Dunn [2] las obtiene aplicando multiplicadores de Lagrange e imponiendo la condición de que las sumas de las probabilidades de un elemento a todos los grupos sea 1.

Ante este desarrollo Flores-Sintas et al. [4] muestra que esta condición es necesaria pero no es suficiente. Es decir, la expresión de las probabilidades dadas por el FCM son magnitudes que sumadas dan 1, pero no tienen por qué ser probabilidades. Y por tanto, la función objetivo a optimizar no es la adecuada y las conclusiones que se obtienen a partir de ella no son correctas.

Además como plasma Flores-Sintas [3] los grados de pertenencia obtenidos con el algoritmo FCM son grados de participación, mientras que en la formulación de la teoría de conjuntos difusos de Zadeh [13], el grado de pertenencia de un punto en el dominio de discurso de un conjunto difuso no depende del grado de pertenencia en otro conjunto difuso definido sobre el mismo universo de discurso. Es decir, si los grados de pertenencia dados por el FCM tuvieran el sentido dado por

Zadeh no se les puede imponer la condición de que en un punto la suma para todos los grupos sea 1. Este es el punto de partida para que Flores-Sintas et al. muestre una concepción diferente de la función objetivo.

El objetivo de este trabajo es dar un marco formal matemático a la función objetivo hallada por Flores-Sintas et al. [3] utilizando las definiciones empleadas por Frank Höppner et al. en [7]. Para nuestro cometido primero haremos un breve resumen de los trabajos de Flores-Sintas [3, 4, 5] para encontrar la función objetivo y darle una interpretación física, y por último desarrollaremos el formalismo matemático.

2. Métrica y Función Acción

Simmons [12] escribe que ciertos hechos *"le sugirieron a Euler que la naturaleza persigue sus fines por medios más eficientes y económicos y que hay simplificaciones ocultas en la base del caos aparente de los fenómenos"*, motivo por el cual Euler creo el cálculo de variaciones y despertó el interés de Hamilton casi un siglo después, formalizando lo que se conoce como *Principio de Hamilton* o *Principio de mínima acción*. Con esta idea, Flores-Sintas [3] pensó en cómo trasladar el *Principio de mínima acción* a la clasificación de un sistema formado por una muestra de patrones. ¿Cuál debería ser el resultado de una traslación del Principio a la detección de grupos en una muestra?: que nos diese el conjunto de grupos más adecuados para la muestra.

Desde luego esta meta es el fin del clustering particional difuso y, como ya hemos dicho, habitualmente se ha considerado el minimizar la suma de los cuadrados de los errores cometidos al estimar que cada grupo está representado por un prototipo como el procedimiento más utilizado. Aquí es donde Flores-Sintas [3] varía el planteamiento del problema, y comienza estudiando la *acción* en el espacio de las características.

La magnitud *acción*, en teoría de variedades, se expresa mediante

$$\int K \sqrt{g} d\Omega,$$

donde K es la curvatura escalar, \sqrt{g} el determinante del tensor métrico en un punto x y $d\Omega = dx^1 dx^2 \dots dx^F$ el elemento de volumen en el espacio de las características. Si $\sqrt{g} d\Omega$ es el elemento de volumen teniendo en cuenta la perturbación representada por g , una métrica diferente de la euclidiana, deformando el espacio y el elemento de

volumen queda deformado por el factor \sqrt{g} . Esta observación es muy pertinente si estamos considerando la posibilidad de utilizar una métrica distinta de la euclidiana. Y ¿por qué?. De nuevo Flores-Sintas [3] muestra la necesidad de tener en cuenta las propiedades locales del espacio de las características, como también reflejar la falta de homogeneidad e isotropía del espacio de las características. Para ello se debería contar con una métrica que lo hiciese.

Si se utiliza la distancia Euclídea estamos suponiendo que el espacio es homogéneo e isótropo. Sin embargo si consideramos que la falta de homogeneidad e isotropía del espacio de las características es debida a la distribución de los patrones de la muestra, la magnitud más utilizada, y, aceptada como la que más información nos da sobre las características espaciales de la muestra, es la matriz de covarianzas en la que se basa la distancia de Mahalanobis.

Por tanto, el principal cometido es encontrar una métrica que tenga en cuenta las propiedades locales del espacio de las características y refleje la falta de homogeneidad e isotropía de este. Apoyándose en la teoría de espacios vectoriales curvos, Flores-Sintas [3], utiliza la matriz de covarianzas de la muestra para encontrar la métrica, representada por g .

Siguiendo el estudio de la *acción*, esta puede expresarse de la siguiente forma tensorial:

$$\int K\sqrt{g} d\Omega = \int \Phi\sqrt{g} d\Omega + \int \frac{\partial\sqrt{g}w^i}{\partial x^i} d\Omega,$$

Al realizar la integración para todo el espacio, el segundo miembro del sumando se anula, y por tanto la *acción* vendrá dada por el primer término de la suma; es decir, la función Φ es característica de la perturbación, o bien en nuestro caso, es característica de la muestra. También la curvatura es característica de la perturbación y está relacionada con la magnitud Φ .

Utilizando los símbolos de Christoffel se obtiene, expresada en normas tensoriales:

$$\Phi = (1 - F) \frac{g_{ik}^{(m)}(x^i - m^i)(x^k - m^k)}{1 + g_{ik}^{(m)}(x^i - m^i)(x^k - m^k)}$$

donde $g_{ik}^{(m)}$ son los elementos del tensor covariante en la media, y m^k las coordenadas del vector que representa la media. Notar que, omitiendo el factor $(1 - F)$, el numerador de la expresión anterior es la distancia de Mahalanobis entre el punto x y la media.

Este es el punto de partida para definir la siguiente función:

$$J_{xm} = \frac{\Phi\sqrt{g(x)}}{1 + d_{xm}^2} \quad (3)$$

El subíndice x muestra que la función está calculada para el elemento x de la muestra, si los sumamos todos

$$\begin{aligned} J_{(m)} &= \sum_{x \in X} J_{xm} \\ &= \frac{\sum_{x \in X} \Phi\sqrt{g(x)}}{(1 - F)} \\ &= \sum_{x \in X} \frac{d_{xm}^2 \sqrt{g(x)}}{1 + d_{xm}^2} \end{aligned} \quad (4)$$

Teorema 2.1 (Flores-Sintas, [3]). *La función $J_{(m)}$ mide la desviación total de la muestra en la hipersuperficie asociada a la métrica.*

Si bien es cierto que la acción está definida para un sistema físico integrando en todo el espacio, la "traslación" a un sistema discreto sería sumando para los patrones de la muestra. Luego, obtener una desviación mínima de los datos a los prototipos, proceso que se realizaría minimizando la función $J_{(m)}$, podría trasladarse como maximizar la acción, justo al contrario que los sistemas físicos. Esto es precisamente lo que nos interesa si queremos detectar grupos en una muestra: tendremos que actuar en contra de la tendencia natural que será la formación de un solo grupo.

La función $J_{(m)}$ esta expresada en (4) considerando la métrica derivada de la matriz de covarianzas en la media. Como la distancia Euclídea es un caso particular de la distancia de Mahalanobis, tomando $g_{(x)} = 1$, la función $J_{(m)}$, utilizando la distancia Euclídea sería:

$$J_{(m)}(Ec) = \sum_{x \in X} \frac{d_{xm}^2}{1 + d_{xm}^2}.$$

Pero al ser la métrica constante en cualquier punto, tiene las mismas propiedades que la media, y por consiguiente, para cualquier punto ν perteneciente al espacio de las características podemos escribir:

$$J_{(\nu)}(Ec) = \sum_{x \in X} \frac{d_{x\nu}^2}{1 + d_{x\nu}^2}.$$

Sin embargo, utilizando la distancia Euclídea implícitamente asumimos que el espacio de las características es homogéneo e isótropo. La existencia de los grupos implica que la homogeneidad e isotropía están rotas [3]. En [6] se propone medir la rotura de la homogeneidad e isotropía mediante un factor r , de modo que la función $J_{(\nu)}(Ec)$ quedaría como

$$J_{(\nu)}(Ec) = \sum_{x \in X} \frac{r^2 d_{x\nu}^2}{1 + r^2 d_{x\nu}^2}. \quad (5)$$

3. Marco formal

A continuación estableceremos las definiciones y resultados necesarios siguiendo a Frank Höppner et al. en [7].

Definición 3.1 (Höppner, [7]). Dado un conjunto $\mathcal{D} \neq \emptyset$ y un conjunto de conjuntos \mathcal{R} con cardinal $|\mathcal{R}| \geq 2$ y tal que exista un conjunto $R \in \mathcal{R}$ con $|R| \geq 2$. Llamaremos a \mathcal{D} un espacio de datos y a \mathcal{R} un espacio resultado. Entonces denominaremos espacio de análisis al conjunto

$$\mathcal{A}(\mathcal{D}, \mathcal{R}) = \{f | f : X \longrightarrow K, \emptyset \neq X \subseteq \mathcal{D}, K \in \mathcal{R}\},$$

donde una aplicación $f : X \longrightarrow K \in \mathcal{A}(\mathcal{D}, \mathcal{R})$ representa al resultado de un análisis de datos mediante una aplicación especial, dado un conjunto de datos $X \subseteq \mathcal{D}$ y un posible resultado $K \in \mathcal{R}$.

Como las soluciones posibles pueden ser numerosas, se introduce una medida indirecta que permita comparar las soluciones.

Definición 3.2 (Höppner, [7]). Dado un espacio de análisis $\mathcal{A}(\mathcal{D}, \mathcal{R})$ llamaremos función objetivo a una aplicación

$$J : \mathcal{A}(\mathcal{D}, \mathcal{R}) \longrightarrow \mathbb{R}$$

De este modo, el valor de $J(f)$ se entiende como una medida de error o calidad, y el propósito será minimizar o maximizar, respectivamente, la función J .

Definición 3.3 (Höppner, [7]). Dado un espacio de análisis $\mathcal{A}(\mathcal{D}, \mathcal{R})$, $X \subseteq \mathcal{D}$, $f : X \longrightarrow K \in \mathcal{A}(\mathcal{D}, \mathcal{R})$, $A_k = f^{-1}(k)$ para $k \in K$. Entonces, diremos que f es una partición si $\{A_k | k \in K\}$ es una partición de X ; es decir,

$$\begin{aligned} \bigcup_{k \in K} A_k &= X, \\ \forall i, j \in K, i \neq j &\Rightarrow A_i \cap A_j = \emptyset, \\ \forall i \in K, \emptyset &\neq A_i \neq X. \end{aligned}$$

Nosotros vamos a trabajar en clustering difuso por tanto introduciremos la definición de conjunto difuso para aplicarlo a las definiciones anteriores.

Definición 3.4 (Zadeh, [13]). Llamaremos función de pertenencia del conjunto X a una aplicación $\mu : X \longrightarrow [0, 1]$. De este modo un conjunto difuso del conjunto X puede representarse por los pares

$$\mu_X = \{(\mu(x)/x) | x \in X\}.$$

El valor $\mu(x)$ describe un grado de pertenencia de x en μ_X .

Definición 3.5. Al conjunto de todas las funciones de pertenencia del conjunto X lo denotaremos por

$$\mathbf{F}(X) = \{\mu | \mu : X \longrightarrow [0, 1]\}$$

Definición 3.6 (Höppner, [7]). Dado un espacio de análisis $\mathcal{A}(\mathcal{D}, \mathcal{R})$, llamaremos espacio de análisis difuso, y lo notaremos como $\mathcal{A}_{dif}(\mathcal{D}, \mathcal{R})$, al conjunto

$$\mathcal{A}_{dif}(\mathcal{D}, \mathcal{R}) = \mathcal{A}(\mathcal{D}, \{\mathbf{F}(K) | K \in \mathcal{R}\}).$$

Definición 3.7 (Höppner, [7]). Dado un espacio de análisis difuso $\mathcal{A}_{dif}(\mathcal{D}, \mathcal{R})$, llamaremos partición probabilística difusa a una aplicación $f : X \longrightarrow \mathbf{F}(K) \in \mathcal{A}_{dif}(\mathcal{D}, \mathcal{R})$, con $X \subseteq \mathcal{D}$, que cumple:

$$\forall x \in X, \sum_{k \in K} f(x)(k) = 1, \quad (6)$$

$$\forall k \in K, \sum_{x \in X} f(x)(k) > 0, \text{ y} \quad (7)$$

Con esta definición diremos que $f(x)(k)$ es la probabilidad de pertenencia del dato $x \in X$ al grupo $k \in K$ relativo a todos los otros grupos.

Así la condición (6) nos dice que la suma de los grados de pertenencia de un dato a cada grupo es 1, condición necesaria para considerar el grado de pertenencia como probabilidad de pertenencia. La segunda condición (7) mantiene que un grupo k no puede ser vacío; es decir, siempre debe existir un dato que tenga probabilidad de pertenencia al grupo.

La condición (6) no siempre se considera, siendo Krishnapuram y Keller ([10], [11]) los que proponen no considerar la restricción (6) y establecen una nueva definición de partición difusa teniendo en cuenta solo la condición (7).

Definición 3.8 (Höppner, [7]). Dado un espacio de análisis difuso $\mathcal{A}_{dif}(\mathcal{D}, \mathcal{R})$, llamaremos partición difusa a una aplicación $f : X \longrightarrow \mathbf{F}(K) \in \mathcal{A}_{dif}(\mathcal{D}, \mathcal{R})$, con $X \subseteq \mathcal{D}$, que cumple: $\forall k \in K$

$$\sum_{x \in X} f(x)(k) > 0$$

Con esta definición diremos que $f(x)(k)$ es el grado de pertenencia del dato $x \in X$ al grupo $k \in K$.

Habitualmente a este tipo de partición se le denomina "Possibilística".

Consideremos ahora la ecuación (5). Esta se puede escribir como

$$J_{(\nu)}(Ec) = \sum_{x \in X} \mu_{x\nu} r^2 d_{x\nu}^2 \quad (8)$$

donde

$$\mu_{x\nu} = \frac{1}{1 + r^2 d_{x\nu}^2}. \quad (9)$$

Del modo que acabamos de definir la función $\mu_{x\nu}$ se ve claramente que puede definirse mediante

$$\nu \in X, \mu_{-\nu} : X \longrightarrow [0, 1], \mu_{-\nu}(x) \longmapsto \mu_{x\nu},$$

lo que le confiere la característica de ser una función de pertenencia del conjunto X , por tanto podemos definir el siguiente conjunto difuso:

$$\bar{\nu} = \{(\mu_{x\nu}/x) \mid x \in X\}.$$

Teorema 3.9. *Dados $X \subseteq \mathbb{R}^F$ y $\mathcal{V} \subset \mathbb{R}^F$ de cardinal finito, entonces el conjunto $\{\bar{\nu} \mid \nu \in \mathcal{V}\}$ es una partición difusa convexa de X .*

Prueba. Como hemos visto $\bar{\nu} = \{(\mu_{x\nu}/x) \mid x \in X\}$ es un conjunto difuso, y además $\bar{\nu} \neq \emptyset \forall \nu \in \mathcal{V}$, por tanto es una partición difusa de X . Solo queda por probar que cada conjunto difuso $\bar{\nu}$ es convexo; es decir, que $\mu_{z\nu} \geq \min\{\mu_{x\nu}, \mu_{y\nu}\}$, para $z = \lambda x + (1-\lambda)y$, $\lambda \in [0, 1] \forall x, y \in \bar{\nu}$. Basta con observar que $d_{x\nu}^2 \leq d_{z\nu}^2 \leq d_{y\nu}^2$ ó $d_{y\nu}^2 \leq d_{z\nu}^2 \leq d_{x\nu}^2$. \square

Teorema 3.10 (Algoritmo Fuzzy Minimals(FM) de Flores-Sintas,[3]). *Sea $p \in \mathbb{N}_{>0}$, $\mathcal{D} = \mathbb{R}^p$, $X = \{x_1, x_2, \dots, x_n\} \subseteq \mathcal{D}$, $\mathcal{C} = \mathbb{R}^p$, $\mathcal{R} = \wp(\mathcal{C})$ y*

$$d_{x\nu} : \mathcal{D} \times \mathcal{C} \longrightarrow \mathbb{R}, (x, \nu) \longmapsto \|x - \nu\|.$$

Entonces existe un conjunto $\mathcal{V} \in \mathcal{R}$, tal que la partición difusa $f : X \longrightarrow \mathbf{F}(\mathcal{V})$, $f(x)(\nu) \longmapsto \mu_{x\nu}$ minimiza la función

$$J = \sum_{x \in X} \sum_{\nu \in \mathcal{V}} \mu_{x\nu} r^2 d_{x\nu}^2$$

siendo

$$\nu = \frac{\sum_{x \in X} \mu_{x\nu}^2 x}{\sum_{x \in X} \mu_{x\nu}^2}.$$

Prueba. Tomando la definición de $\mu_{x\nu}$ dada en (9), la función objetivo J podemos ponerla como

$$J = \sum_{x \in X} \sum_{\nu \in \mathcal{V}} (1 - \mu_{x\nu}).$$

Supongamos que efectivamente existe \mathcal{V} . Para saber si la función tiene un mínimo en $\nu \in \mathcal{V}$ calcu-

lamos la derivada parcial respecto de ν

$$\begin{aligned} \frac{\partial J}{\partial \nu} &= \frac{\partial}{\partial \nu} \sum_{x \in X} \sum_{\nu \in \mathcal{V}} (1 - \mu_{x\nu}) \\ &= - \sum_{x \in X} \frac{\partial \mu_{x\nu}}{\partial \nu} \\ &= 2 \sum_{x \in X} \mu_{x\nu}^2 (x - \nu) \end{aligned}$$

De aquí es fácil deducir que

$$\begin{aligned} \frac{\partial J}{\partial \nu} = 0 &\iff \sum_{x \in X} \mu_{x\nu}^2 (x - \nu) = 0 \\ &\iff \nu = \frac{\sum_{x \in X} \mu_{x\nu}^2 x}{\sum_{x \in X} \mu_{x\nu}^2}. \end{aligned}$$

La existencia del conjunto \mathcal{V} se puede ver en Flores-Sintas [3] quién construye un algoritmo que nos da los elementos de \mathcal{V} .

Por último sólo quedaría por probar que es una partición difusa. Por la definición $\mu_{x\nu}$ y la manera de obtener los elementos $\nu \in \mathcal{V}$ siempre existirá algún dato $x \in X$ lo suficientemente cerca de ν para que $\mu_{x\nu} > 0$, y por tanto $\forall \nu \in \mathcal{V}$, $\sum_{x \in X} \mu_{x\nu} > 0$. \square

La partición difusa que hemos creado no es probabilística pero podemos hacer que lo sea.

Teorema 3.11. *Si $\mathcal{V} = \{\nu_1, \nu_2, \dots, \nu_c\}$ es el conjunto obtenido en el Teorema 3.10, entonces la aplicación $f : X \longrightarrow \mathbf{F}(\mathcal{V})$,*

$$f(x)(\nu_i) \longmapsto u_{x\nu_i} = \frac{\mu_{x\nu_i} \sqrt{g_{\nu_i}} \sum_{y \in X} \mu_{y\nu_i}}{\sum_{j=1}^c \mu_{x\nu_j} \sqrt{g_{\nu_j}} \sum_{y \in X} \mu_{y\nu_j}} \quad (10)$$

donde g_{ν_i} es el determinante de la inversa de la matriz de covarianzas en ν_j , es una partición probabilística difusa.

Prueba. Las condiciones (6) y (7) se deducen fácilmente de las definiciones de u_{xi} y $\mu_{x\nu_i}$ respectivamente. \square

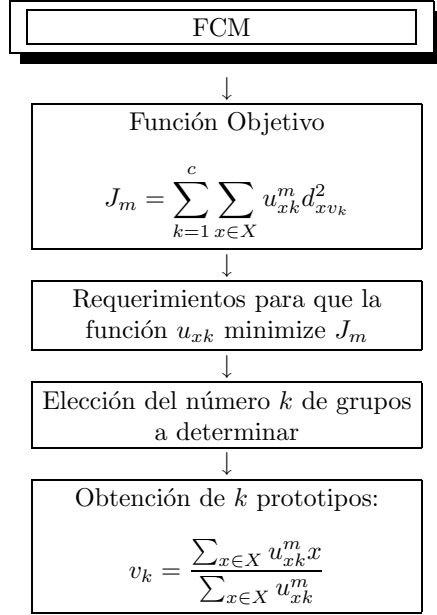


Figura 1. Esquema de funcionamiento del FCM clásico

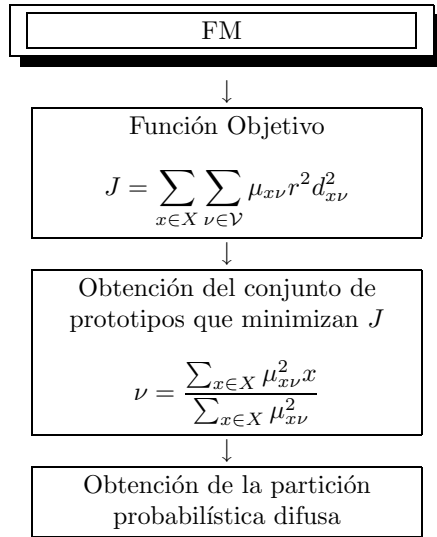


Figura 2. Esquema de funcionamiento FM

El algoritmo FM se sale de la órbita de los métodos derivados del FCM puesto que no necesitamos considerar un número de prototipos determinado, pero mantiene la arquitectura del FCM que implica una minimización de una función objetivo y una siguiente optimización del proceso resolutorio que dé la mejor partición: *Alternating Optimization (AO)*.

Vemos que el algoritmo FM trabaja sólo con la distancia Euclídea. No obstante, como refleja Flores-Sintas, la posibilidad de detectar los grupos sin conocimiento previo sólo es posible si la métrica es la misma para toda la muestra. En este sentido, si suponemos que la métrica es constante,

por ejemplo que sea la distancia de Mahalanobis utilizada como norma, también es posible trabajar con el algoritmo FM. Sin embargo, en el caso de la distancia de Mahalanobis, es más interesante obtener una métrica para cada grupo. Además a veces podemos necesitar elegir nosotros el número de grupos. En tal caso la alternativa del FCM es idónea.

Como habíamos dicho la función $J_{(m)}$ está expresada en (4) considerando la métrica derivada de la matriz de covarianzas en la media, pero es más interesante obtener una métrica para cada grupo. Sea $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^F$ una muestra que queremos dividir en c grupos difusos. Sea $\mathcal{V} = \{\nu_1, \nu_2, \dots, \nu_c\} \subset \mathbb{R}^F$ el conjunto de los prototipos de los grupos que queremos obtener. Representemos por g_k al determinante de la inversa de la matriz de covarianzas difusa del grupo k . En [5] Flores-Sintas propone que la métrica definida por la matriz de covarianzas difusa del grupo k se comporta igual que la métrica considerando ν_k como si fuese la media del grupo k . Entonces la función que nos da la desviación para el grupo k es:

$$J_k = \sum_{x \in X} \mu_{xk}^{3/2} d_{xk}^2 \sqrt{g_k} \quad (11)$$

donde, d_{xk}^2 es distancia de Mahalanobis entre x y ν_k , y $\mu_{xk} = \frac{1}{1+d_{xk}^2}$. La probabilidad de que un patrón x pertenezca al grupo k será proporcional a la densidad del grupo k en x . En consecuencia, la probabilidad es:

$$u_{xk} = \frac{\mu_{xk}^{3/2} \sqrt{g_k}}{\sum_{j=1}^c \mu_{xj}^{3/2} \sqrt{g_j}}. \quad (12)$$

Teorema 3.12. Sea $p \in \mathbb{N}_{>0}$, $\mathcal{D} = \mathbb{R}^p$, $X = \{x_1, x_2, \dots, x_n\} \subseteq \mathcal{D}$, $\mathcal{C} = \mathbb{R}^p \times \{g \in \mathbb{R}^{p \times p} | g \text{ es simétrica y definida positiva}\}$, $c \in \mathbb{N}$, $\mathcal{R} = \wp(\mathcal{C})$, $d_{xk} = \{d_{xk}^{(Mh)}, d_{xk}^{(Ec)}\}$ donde

$$\begin{aligned} d_{xk}^{(Mh)} : \mathcal{D} \times \mathcal{C} &\longrightarrow \mathbb{R}, \\ (x, (\nu_k, g_k)) &\longmapsto (x - \nu_k)^\top g_k (x - \nu_k), \\ d_{xk}^{(Ec)} : \mathcal{D} \times \mathbb{R}^p &\longrightarrow \mathbb{R}, \\ (x, \nu_k) &\longmapsto \|x - \nu_k\|. \end{aligned}$$

Si la función

$$J = \sum_{x \in X} \sum_{k=1}^c u_{xk}^2 d_{xk}^2,$$

siendo u_{xk} la función de pertenencia de la partición probabilística difusa $\{\{(u_{xk}/x) | x \in X\} | \forall u_{xk}, 1 \leq k \leq c\}$ dada en (12) por $d^{(Mh)}$

y (10) para $d_{xk}^{(Ec)}$, tiene un mínimo, entonces

$$\nu_k = \frac{\sum_{x \in X} u_{xk}^2 x}{\sum_{x \in X} u_{xk}^2}.$$

Prueba. La prueba de este resultado sigue los mismos pasos que la dada por Frank Höppner et al. en [7] para obtener los prototipos del FCM. \square

4. Conclusión

Como habíamos pretendido hemos expresado mediante teoremas algunos de los enunciados formulados por Flores-Sintas, probándose la convexidad de la nueva función de pertenencia, cuestión que mostraremos importante en futuros trabajos, y que justifica el mejor cumplimiento de las consideraciones de Zadeh de esta función de pertenencia que la función de pertenencia presente en el FCM clásico. Además la nueva función objetivo tiene un significado físico concreto, y no necesita de la arbitrariedad para la elección del parámetro de "fuzzificación".

Referencias

- [1] J.C. Bezdek. A convergence theorem for the fuzzy isodata clustering algorithms. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-2:1–8, 1980.
- [2] J. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well separated cluster. *J. Cybern.*, 3:32–57, 1974.
- [3] A. Flores-Sintas, J.M. Cadenas, and F. Martin. A local geometrical application to fuzzy clustering. *Fuzzy Sets and Systems*, 100:245–256, 1998.
- [4] A. Flores-Sintas, J.M. Cadenas, and F. Martin. Membership functions in the fuzzy c-means algorithm. *Fuzzy Sets and Systems*, 101:49–58, 1999.
- [5] A. Flores-Sintas, J.M. Cadenas, and F. Martin. Partition validity and defuzzification. *Fuzzy Sets and Systems*, 112:433–447, 2000.
- [6] A. Flores-Sintas, J.M. Cadenas, and F. Martin. Detecting homogeneous groups in clustering using the euclidean distance. *Fuzzy Sets and Systems*, 120:213–225, 2001.
- [7] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis: Methods for classification, data analysis, and image recognition*. John Wiley & Sons, 2000.
- [8] M.S. Kamel and S.Z. Selim. A relaxation approach to the fuzzy clustering problem. *Fuzzy Sets and Systems*, 61:177–188, 1994.
- [9] R. Krishnapuram. Generation of membership functions via possibilistic clustering. In *Procc. IEEE World Congress on Computational Intelligence*, pages 902–908, Orlando (U.S.A.), 1994.
- [10] R. Krishnapuram and J.M. Keller. A possibilistic approach to clustering. *IEEE Transactions On Fuzzy Systems*, 1(2):98–110, May 1993.
- [11] R. Krishnapuram and J.M. Keller. The possibilistic c-means algorithm: insights and recommendations. *IEEE Transactions On Fuzzy Systems*, 100:245–256, August 1996.
- [12] George F. Simmons. *Ecuaciones diferenciales con aplicaciones y notas históricas*. McGraw-Hill, 1977.
- [13] L.A. Zadeh. Fuzzy sets. *Information and Control*, pages 338–353, 1965.