
ANÁLISIS DE LOS ALGORITMOS DE AGRUPAMIENTO BORROSO PARA DETECTAR ASIMETRÍA DE INFORMACIÓN.

EDUARDO CANO PLATA¹, CARLOS CANO PLATA²

RESUMEN

Este documento muestra una técnica ideal para la toma de decisiones empleando la lógica borrosa con el método de agrupamiento de sus siglas en inglés "Fuzzy clustering method" (FCM). El mismo se empleó para discriminar variables relevantes en la prestación de servicios de salud, teniendo como referencia información derivada de una encuesta aplicada directamente a los usuarios del servicio, con el propósito de encontrar asimetría de información.

Los modelos borrosos permiten la discriminación de parámetros sumamente complejos para su clasificación, tales como los que podrían determinar la asimetría de información. Utilizando puntos medios borrosos se generan los grupos que permiten evaluar la correlación paciente médico. El conocimiento previo de las evaluaciones estadísticas permite caracterizar un modelo borroso. La información recolectada se integró en el algoritmo, quien la agrupa, utilizando un prototipo de patrón (el sistema de entrenamiento). Los grupos se construyen empleando una función de decisión. La característica para la toma de decisiones en el evento de pensar una visita al médico, se encuentra a través de un análisis del componente principal utilizando el procedimiento FCM. El algoritmo de agrupamiento mostró ser una manera confiable de identificar la asimetría existente en la relación paciente medico.

Palabras claves: Método de agrupamiento borroso, servicio de salud, asimetría de información.
JEL: D230, O120, I110.

ABSTRACT

This document characterizes a decision-making method combining Fuzzy logic with the clusters method known as the "Fuzzy clustering method " (FCM). This method is used to discriminate between relevant variables in health services, where the reference information is derived from a survey directly applied to users in order to detect asymmetric information.

Fuzzy models have performed well in Classification theory, especially in the identification of asymmetry of information because they can be used to discriminate extremely complex parameters. Using average fuzzy points (the crips), groups are generated to evaluate the patient-doctor correlation. Previous knowledge of the statistical evaluations can be characterized using a fuzzy model.

1 Doctor de la Universidad de Buenos Aires, área Ingeniería. Director Grupo de Redes de Distribución y Potencia Departamento de Ingeniería Eléctrica, Electrónica y Computación. Universidad Nacional Sede Manizales. Campus La Nubia. A.A. 127 Tel/Fax: +57 68 743032 Manizales – Colombia E-mail: ecano@ieec.org

2 Universidad de Bogotá Jorge Tadeo Lozano Grupos: AdGeo (Universidad de Bogotá Jorge Tadeo Lozano Categoría Reconocido Conciencias); Competitividad Empresarial y Gestión Tecnológica. (Universidad Nacional de Colombia Categoría C) Profesional: Administrador de Empresas Universidad Nacional de Colombia. Especialista en Derecho Empresarial (UNAB). Magíster en Administración de Empresas Universidad Nacional de Colombia. Asesor en Seguridad Social.

The clusters are constructed using a decision function, where the decision to be captured is thinking a visit to the doctor. A principal component analysis is performed using the procedure FCM. The clustering algorithm showed that it is a reliable way of identifying the existing asymmetry in the patient-doctor relation.

Key words: Fuzzy clustering Method, service of health, asymmetric information

I. INTRODUCCIÓN

En este documento se pretende mostrar un método para resolver problemas complejos de toma de decisiones con múltiples criterios. Los métodos “puros” por si solos han resultado ineficientes; en éstos casos se requiere del uso de métodos híbridos, que permitan combinar los métodos jerárquicos con otras técnicas, con el fin de obtener soluciones más satisfactorias para los involucrados en el proceso. En la ciencia los esfuerzos se dirigen en dos sentidos: el conocimiento objetivo, esto quiere decir hacia los hechos, y lo que queda únicamente en la mente humana, llamados entes ideales (Gil Aluja, 2002). Para el caso se ha tomado como objeto de análisis el servicio de salud y se pretende discriminar variables relevantes en la prestación de servicio en el cafetero colombiano, con el objetivo de evaluar la decisión de la toma del servicio por parte del paciente y la asimetría en la relación paciente medico.

De la teoría de señales en los sistemas dinámicos, se tiene que extraer la información (análisis) proveniente de una encuesta no selectiva que es fundamental para la evaluación de parámetros que de una u otra forma ayudarán a la toma de decisiones. Estas características o parámetros de la señal, deben ser resaltados y además se deben ordenar (Duda et al, 1989).

Existen diferentes herramientas econométricas para este propósito: mínimos cuadrados (LS), métodos no paramétricos (NP), método binomial negativo (Negbin-2) y el método generalizado de los momentos (GMM), se conoce ampliamente como técnica de clasificación (Duda et al, 2000).

La técnica de agrupamiento posee la estrategia de generar subconjuntos en un espacio n- dimensional de las características de una señal, por lo que consideramos que este método es el apropiado para identificar variables multicriterio (Duda et al, 2000).

Por ejemplo, el método de agrupamiento ha sido ampliamente utilizado en la discriminación de patrones de bebidas y licores en la industria de alimentos (Kobayashi & Aishima, 1995). Shao y Wu (2005) describe el problema de regresión para determinar el número de grupos y propone un criterio para seleccionar el número regresión en un hiperplano con probabilidad de seleccionar todas las secuencias de clasificación, cuando el número de observaciones tiende a infinito. David Giles y Rober Draeseke (2001) han utilizado un arreglo de cinco factores para modelar la relación edad contra ingresos, empleando econometría difusa con el método de de agrupamiento. Por otra parte, Lin et al (2008) han empleado la técnica borrosa para el análisis del contrato de gestión debido a la existencia de asimetría de la información.

Este artículo utilizará los algoritmos Fuzzy c-means (FCM) y el de Gustafson-Kessel (GK) (Kim et al, 2004) para la clasificación y selección de los factores que inciden en la toma de decisiones por parte de los pacientes en el momento de asistir al médico.

II. AGRUPAMIENTO BORROSO

El objetivo de análisis del agrupamiento es la clasificación de objetos según las similitudes entre ellos, así como la organización de los datos en subconjuntos. Un grupo es un subconjunto de objetos que son muy similares los unos entre los otros. En los espacios métricos se define a menudo la similitud por medio de la distancia, basada en la longitud de un vector a un punto del conjunto de datos que formará el grupo. Los centros de estos subconjuntos no tienen una relación inicial (no se conocen de antemano), lo que hace que el algoritmo localice su inicio haciendo una partición de estos datos de manera aleatoria. Por consiguiente, la técnica de agrupamiento busca subconjuntos de los datos, sin que ellos tengan conocimiento a priori para identificar las clases. Los puntos de inicio (centros, prototipos o crips) deben ser vectores de la misma dimensión de los otros datos; pero ellos se definen como objetos geométricos de alto nivel, tales como espacios funcionales, lineales o no-lineales (Edwards, 1995). Dado que los grupos pueden ser formalmente vistos como un subconjunto de datos, un posible método de clasificación es basarse en la teoría clásica de conjuntos y decir entonces que un elemento está o no incluido dentro de un grupo. Los métodos FCM (Gulley Jang, 2000) tienen la ventaja sobre la lógica de conjuntos en el hecho de decir que un elemento puede estar incluido en varios grupos simultáneamente, con diferentes grados de pertenencia. El conjunto de datos Z es entonces dividido en c subconjuntos borrosos.

En muchas situaciones reales, el grupo borroso es mucho más natural que la teoría de conjuntos convencional, puesto que los objetos que se encuentran en las fronteras de estos grupos no necesariamente pueden ser forzados a pertenecer a uno de ellos. Por esto, este problema de la frontera puede ser resuelto asignando un grado de pertenencia en el intervalo $[0, 1]$, que indicará una pertenencia parcial.

En este artículo se considera la cuantificación de datos a través del Agrupamiento Borroso.

Los datos que se utilizan son un ordenamiento resultado de la observación de una encuesta. En términos generales, cada observación consiste de la medida de n variables, que se agrupan en un vector n - dimensional, este será la columna $z_k = [z_{k1}, \dots, z_{kn}]^T$, $z_k \in \mathbb{R}^n$. Así un conjunto de N observaciones se denotará por $Z_k = \{z_k / k = 1, 2, \dots, N\}$ que puede ser representado por una matriz $n \times N$:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & z_{22} & \cdots & z_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nN} \end{bmatrix} \quad (1)$$

En la terminología de reconocimiento de patrones, los objetos v_i son los centros del grupo a ser caracterizado (Duda et al, 2000), las columnas de Z son llamados patrones y las filas de Z son llamadas atributos; a Z se le conoce con el nombre de la matriz de patrones.

El objetivo del agrupamiento es dividir el conjunto de datos Z en c grupos. La matriz de $c \times N$, $U = [\mu_{ik}]$ representa la partición borrosa de los elementos que satisfacen las siguientes condiciones:

$$\mu_{ik} \in [0, 1] \quad 1 \leq i \leq c, 1 \leq k \leq N \quad (2)$$

$$\sum_{i=1}^c \mu_{ik} = 1, 1 \leq k \leq N \quad (3)$$

$$0 < \sum_{k=1}^N \mu_{ik} < N, 1 \leq i \leq c \quad (4)$$

Donde: c es el número de los grupos borrosos y μ_{ik} denota el grado de pertenencia, de $z_k = [z_{k1}, \dots, z_{kn}]^T$ - la observación del i -ésimo grupo estará entre $1 \leq i \leq c$ siendo c el centro del subconjunto.

El objetivo del modelo de FCM (Gulley Jang, 2000) es minimizar la suma media cuadrática ponderada de las distancias entre los puntos z_k y el centro del grupos v_i , las distancias $D_{i,k}^2$, son ponderadas con el valor de pertenencia $\mu_{i,k}$.

Por consiguiente, la función objetivo es:

$$J(Z, U, V) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{i,k}^2 \quad (5)$$

Donde $U = [\mu_{ik}]$ es la matriz de partición borrosa de Z , $V = [v_1, v_2, \dots, v_c]$ es el vector de centros, críps o patrones de los grupos, $m \in (1, \infty)$ es un exponente de ponderación que determina el grado de borrosidad del grupo; a menudo se toma $m = 2$. $D_{i,k}^2$ puede ser determinado eligiendo una norma apropiada, por ejemplo la A-norm (Edwards, 1995).

$$D_{ik}^2 = \|z_{k_i} - v_i\|_A \sqrt{(z_k - v_i)^T A (z_k - v_i)} \quad (6)$$

La minimización del funcional c-means (ecuación 8) representa un problema de optimización no-lineal que puede resolverse usando cualquiera de los métodos disponibles (Díez y Navarro, 2004). El método más popular; sin embargo, es la optimización alterna (AO), conocido como FCM-AO (de sus siglas en inglés "Fuzzy c-means alternating optimization"), el cual se resume en la Tabla N° 1.

Los puntos utilizados como críps, en el FCM, producen los clusters esféricos (correspondientes a la A-norm). Pueden obtenerse diferentes formas del cluster con la

selección de normas diferentes como lo sugerido en el algoritmo de Gustavson-Kessel (Kim et al, 2004), o con los tipos diferentes de partición; por ejemplo, las variedades funcionales lineales (FCV) (Diez y Navarro, 2004), dónde los grupos son subespacios lineales del espacio de rango. Una variedad lineal r-dimensional es definida por el vector y su dirección. En este caso, la distancia entre los datos y el cluster i-ésimo es:

$$D_{ik}^2 = \sqrt{\|z_k - v_i\|^2 - \sum_{j=1}^r ((z_k - v_i)^T A s_{ij})^2} \tag{7}$$

La correspondiente variedad funcional “c-means borrosa” para la optimización alterna (FCV-AO) planteada para determinar los centros v_i en el paso 1 de la Tabla N°1, computa las direcciones de s_{ij} como el vector propio unitario del valor propio r más grande de la matriz borrosa:

$$S_{iA} = A^2 \left[\sum_{k=1}^N \mu_{ik} (z_k - v_i)(z_k - v_i)^T \right] A^2 \tag{8}$$

Si el $r=1$, esto produce los algoritmos c-líneas (FCL) y algoritmo de FCL-AO.

La partición automática de la matriz se da por el algoritmo de Gustafson-Kessel, que es una manera natural de determinar los radios relacionándolos al tamaño de los grupos. Esto puede lograrse considerando la matriz de covarianza:

$$P_i = \frac{\sum_{k=1}^N \mu_{ik}^c (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^N \mu_{ik}^m} \tag{9}$$

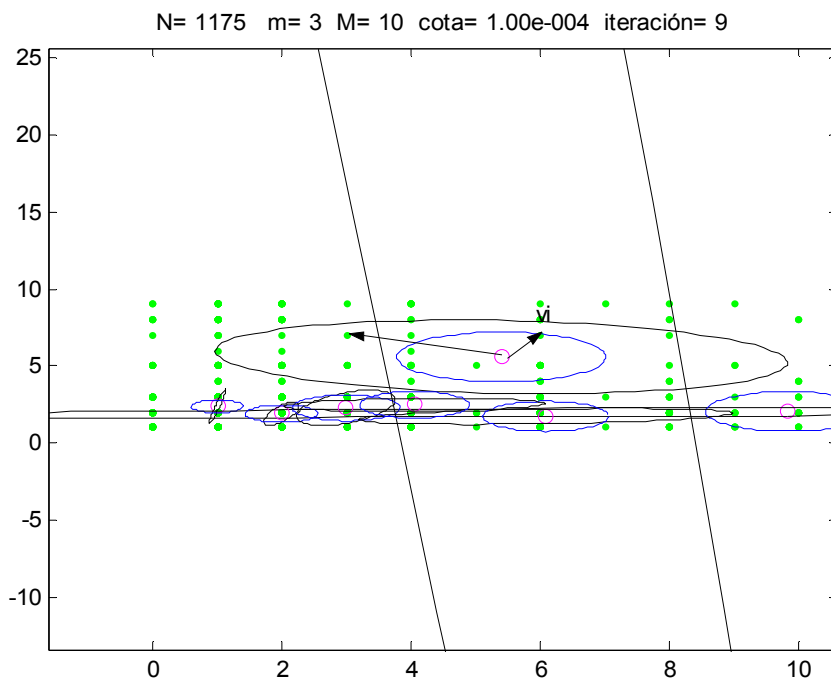
El determinante $|P_i|$ de la matriz de covarianza del grupo entrega el volumen del subconjunto. Puesto que P_i es una matriz definida positiva y simétrica, puede descomponerse tal que $P_i = Q_i \Lambda_i Q_i^T$ donde Q_i es orthonormal y Λ_i es diagonal con elementos no ceros $\lambda_{i1}, \dots, \lambda_{im}$.

Permitiendo que el volumen de los grupos se extienda una distancia de $\sqrt{\lambda_{ij}}$, $j = 1, 2, \dots, n$ a lo largo de cada vector propio q_{ij} . En el caso unidimensional, esta opción implica que el “crisp” o centro del grupo se extienda una desviación normal del centro del subconjunto. En el caso multi-dimensional, el tamaño del radio en cada dirección es determinado midiendo las distancias a lo largo de las coordenadas transformadas según

$$\sqrt{\Lambda_i} Q_i^T A_i Q_i \sqrt{\Lambda_i}, \tag{10}$$

Donde $\sqrt{\Lambda_i}$ representa una matriz cuyos elementos son iguales a la raíz cuadrada de los elementos Λ_i . Cuando se induce una norma diferente dada por la matriz de la covarianza, se obtendrán distintos valores para el radio. En este caso, se puede determinar con un valor inicial promedio, como se discutirá mas adelante. La forma del volumen de los centros del grupo o prototipos, es igual a la forma del grupo inducido por la distancia métrica. Cuando se emplean distancias euclidianas para el algoritmo FCM, los crips y sus hiperformas se muestran en la Figura N°1.

FIGURA N°1
GEOMETRÍA DE LAS DISTANCIAS



Fuente: Elaboración propia.

TABLA N°1
ALGORITMO DE AGRUPAMIENTO DIFUSO

<p>La inicialización: Dado un conjunto de datos Z, seleccione el número de grupos c, el exponente ponderado m y el criterio de tolerancia $\varepsilon \leq 0.001$ e inicialice la partición de la matriz al azar.</p>
<p>Paso 1: Determine los centros de los grupos:</p> $v_i^{(1)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m z_k}{\sum_{k=1}^N (\mu_{ik}^{(1-1)})^m}, 1 \leq i \leq c \tag{11}$
<p>Paso 2.: calcule las distancias</p> $D_{ik}^2 = \ z_k - v_i\ ^2 = \sqrt{(z_k - v_i)^T A(z_k - v_i)}, 1 \leq i \leq c, 1 \leq k \leq N \tag{12}$ <p>Paso 3.: Reorganice nuevamente la matriz de partición con las siguientes reglas:</p> <p>Si $D_{ik} > 0$ para $1 \leq i \leq c, 1 \leq k \leq N$,</p> $\mu_{ik}^{(1)} = \frac{1}{\sum \left(\frac{D_{IK}}{D_{JK}} \right)^{2/m-1}} \tag{13}$ <p>entonces:</p> $\mu_{ik}^{(1)} = 0 \tag{14}$ <p>De otra forma: $\mu_{ik}^{(1)} = 0$ (15)</p> <p>Hasta : $U^{(l)} - U^{(l-1)} < \varepsilon$ (16)</p>
<p>Repita para i= 2...hasta l</p>

Fuente: Elaboración propia.

II.1. Experimento

Se formuló un arreglo a partir de una encuesta. La encuesta fue construida utilizando un muestreo aleatorio simple para poblaciones infinitas. El grupo encuestado de la población permitió detectar información asimétrica. Los factores que explican la decisión de visitar al médico general, por parte de los pacientes también se derivó del método. Se hicieron cuatro preguntas, aplicadas a una distribución normal dentro de la población escogida. Las preguntas fueron direccionadas a diferentes tipos de población de acuerdo

a su discriminación socioeconómica en el eje cafetero colombiano, estratos: I, II, III, IV, V, VI respectivamente. Las variables socioeconómicas fueron determinadas de manera experimental utilizando técnicas borrosas; los datos extraídos del nivel tres, se analizaron para ver su flujo en cuanto a variables endógenas y variables inobservables y la interferencia del muestreo. Se construyó una tarjeta de datos manualmente para entregarlo a un programa de computador desarrollado en MATLAB 6.1 (MatLab 6.1 , 2000), cuyo objetivo es medir el cambio en la distribución. Como se ha dicho en función de detectar asimetría, la encuesta también fue aplicada a un grupo de médicos.

II.1.1. Metodología

A continuación hacemos una descripción que utiliza la lógica borrosa como base para diseñar una herramienta automática que permite valorar el grado de asimetría en la relación médico paciente (Figura N° 2). Esta herramienta es el resultado de una metodología de diseño utilizada tradicionalmente en la ingeniería (Del Brio y Sanz, 2007). Indica como objetivo primordial valorar la geometría de los grupos que será asociada al grado de asimetría de la información. La lingüística que se presenta a través de la interpretación de la información, son los datos de entrada al sistema experto-borroso objeto resultado del diseño (Sáez, Cipriano y Andrzej, 2002).

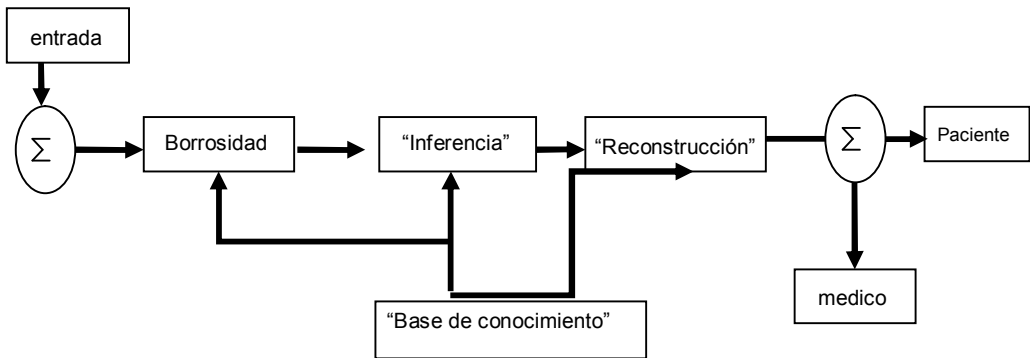
Esta lingüística se obtiene mediante la aplicación de una encuesta cuyo registro técnico era el siguiente:

1. Universo: Población afiliada a la seguridad del Estado de Caldas, Risaralda y Quindío.
2. Unidad de muestras: la seguridad del Estado de la región colombiana de café.
3. Fechas: 30 enero de 2006.
4. Región: Caldas, Risaralda y Quindío.
5. Tipo de muestra: aleatoria simples de poblaciones infinitas.
6. Técnica de recolección de datos: entrevistas.
7. Tamaño de la muestra: 450.
8. Entrenador de fechas: Noviembre de 2005: 25.
9. Número de preguntas: 20.

El diseño del sistema experto se dividió en las siguientes etapas.

1. Etapa de interpretación lingüística: Entrada de datos.
2. Etapa de borrosidad.
3. Inferencia: motor heurístico a través de algunas normas de decisión (paso 3 del algoritmo presentado en la Tabla N°1) base de conocimiento aparta a la clasificación.
4. Reconstrucción y medida de las distancias.
5. Relación de salida entra distancias del paciente al médico.

FIGURA N°2
DISEÑO DEL SISTEMA EXPERTO, ESQUEMA GENERAL



Fuente: Elaboración propia.

II.2. Relación entre el método de agrupamiento borroso y el arreglo formulado por la encuesta

Al detectar los datos entregados por la encuesta, el algoritmo de agrupamiento borroso trata de detectar información de las señales que entrega cada pregunta. Lo que se denomina relación del captador y acondicionamiento de la señal en la jerga de la ingeniería.

La visita al medico o la decisión que tomó cada individuo encuestado (medico o paciente), fue analizada a través de cuatro preguntas simultáneas. Estas preguntas tuvieron diferentes respuesta en relación con la visita al servicio de salud. El rango se obtuvo entre 0,1 a 1 normalizado, utilizando el criterio de entropía y es dependiente del número de preguntas de la encuesta (Lozano y Fuentes,2001) o lo que es lo mismo, pasarlo de normalizado a porcentual (0,01% al 100%), esto es: se tomó el resultado natural de cada respuesta, sumando cada componente del vector y dividiendo cada componente por este resultado. Entonces, obtenemos las valoraciones normalizadas. Ellas fueron agrupadas en cuatro vectores columnas, en relación a un solo experimento, la relación de las frecuencias medidas es una característica y puede ser utilizada para discriminar factores de decisión respecto de ir al médico. De esta forma, cuando evaluamos el k-ésimo experimento, el programa computacional leerá cuatro valores de frecuencia:

$$z_i = [H_i, I_i, G_i, S_i]^T, z_i \in \mathfrak{R}^4 \tag{17}$$

Donde

- H = Es el estado de salud.
- I = Cubrimiento de seguro individual.
- G = Demográfica.
- S = Estatus socio económico.

Estas variables no necesitan ser caracterizadas, pues surgen de manera natural de la relación lingüística de la encuesta.

II.3. Resultados y discusión

II.3.1. Identificación de 4 factores utilizando el método FCM

El objetivo de utilizar este método es el reconocimiento de cuatro factores en las señales resultantes de la encuesta (Figura N°3). Por ello, el número de factores escogidos es $c = 4$ y el número de vectores Z_i es igual al número de sensores.

El número de observaciones es de 250 por cada factor, así que el tamaño de la matriz de patrones Z es $(250 \times 4 \times 4)$. La ponderación del exponente m fue 2 con la idea de obtener una buena partición. Si se piensa hacer tender m hacia infinito entonces en V_{250i} , i será el valor promedio del grupo donde el índice 250 denota el factor de distribución m ; es decir, cuando $m \rightarrow \infty$ la partición irá al máximo ($C_{IK} / 1 = \mu$). El factor de sensibilidad será ϵ , se le asignó un valor de 10^{-5} . El FCM se realizó utilizando MATLAB 6.1 y el Fuzzy logic Toolbox (Gulley y Jang, 2000)

TABLA N° 2
FACTORES EXPLICATIVOS DE LA TOMA DE DECISIONES EN CUATRO EXPERIMENTOS (CRIPS O CENTROS)

ACTORES	EXPERIMENTO	H	I	G	S
Médico	1	50. 4455	2.3455	8.1 238	4.2 211
Paciente	2	5.5 963	1.7797	1.5 069	2.3 440
Médico	3	6.5 966	1.9735	2.0 749	2.7 313
Paciente	4	5.6 108	1.8965	2.2 286	2.5 390

Fuente: Elaboración propia.

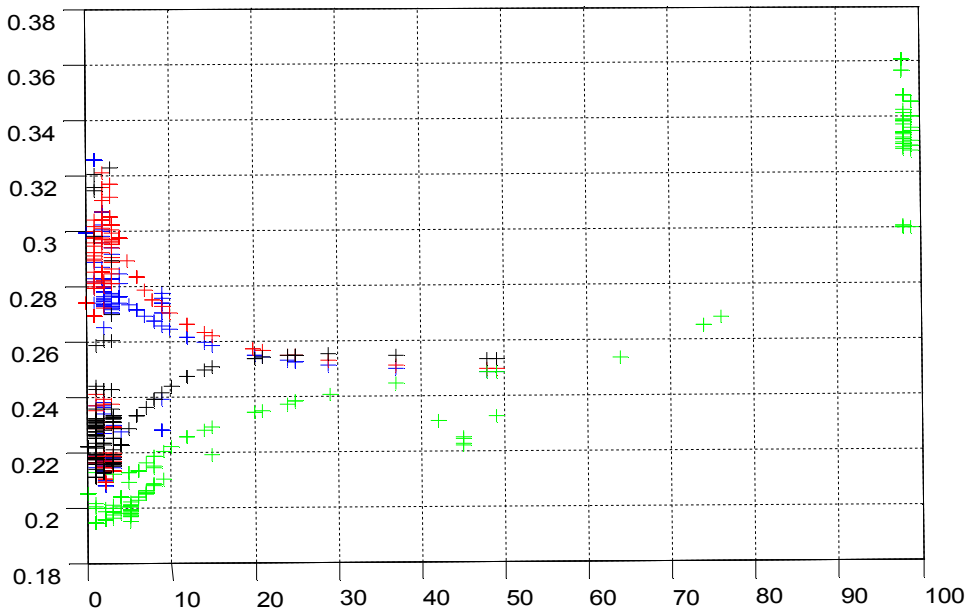
Utilizando el algoritmo propuesto se determinaron los centros que están resumidos en la Tabla N°2. La eficiencia de este algoritmo depende en gran medida de la matriz de inicialización, una función objetivo pequeña se logra para $m=20$, la clasificación no ha observado falla para el análisis del seguro en salud, clase demográfica y estatus de salud; se discriminó bastante bien. Las funciones de pertenencia para $m=20$ se muestran en la Figura N°2.

II.3.2. Identificación de los cuatro factores utilizando el algoritmo Gustafson-Kessel

El método Gustafson-Kessel fue también empleado para identificar estas variables, pero falló cuando trataba de identificar variables entremezcladas. Para resolver este problema se utilizó conocimiento a priori y luego se aplicó el método de Gustafson-Kessel; debido a la ecuación fundamental de VERA, las visitas al médico pueden ser escogidas a través de una ecuación dinámica positiva (Vera,2001):

FIGURA N°3 DISCRIMINACIÓN FCM

Discriminación FCM, función de pertenencia, m=20



Fuente: Elaboración propia.

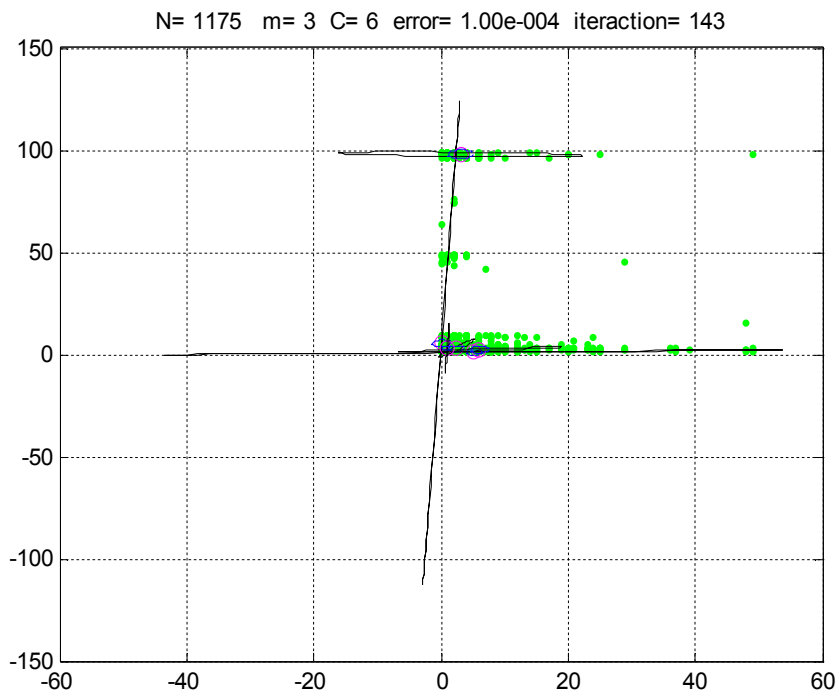
$$v_i = \exp(\beta G_i + \phi I_i) + u_i \tag{18}$$

Donde la función exponencial es empleada para identificar su positividad y u_k se refiere al error del funcional econométrico que también incluye variables no observables.

Teniendo este conocimiento previo, los grupos al ser de tipo exponencial, con radio igual a 1 y partiendo del origen, (es decir, $\beta=0$ y $\phi=0$), en este caso s_{i1} tiene que ser igual a:

$$s_{i1} = \frac{v_i}{|v_i|} \tag{19}$$

FIGURA N°4
MÉTODO DE DISCRIMINACIÓN CON EL ALGORITMO GUSTAFSON-KESSEL

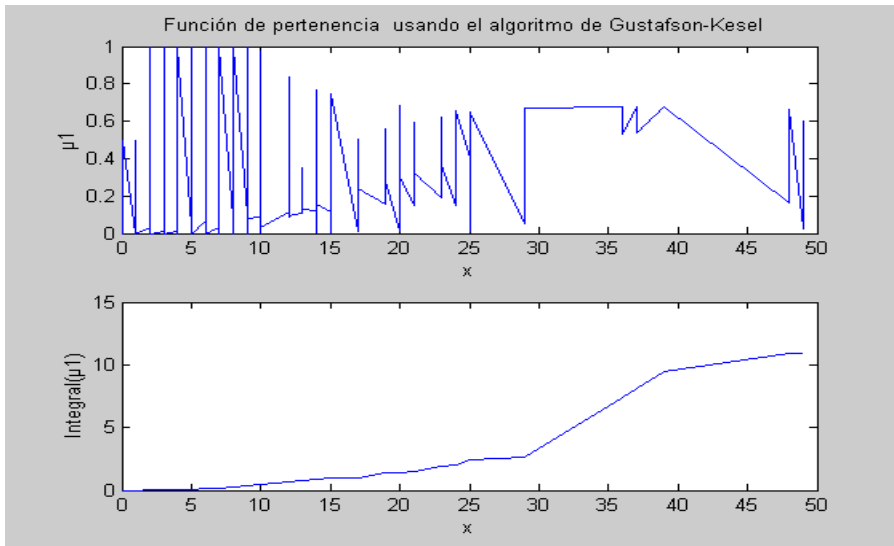


Fuente: Elaboración propia.

Lo que nos muestra un valor propio unitario en la matriz difusa, ecuación (8). Utilizando la ecuación (10) se puede resolver de manera apropiada, incorporando este conocimiento a priori. De esta manera hacer la partición o determinar los grupos será rápido, pues no se tiene que resolver el problema de valores propios de la ecuación (8).

Se hicieron cuatro preguntas dentro de una distribución a cuatro experimentos a poblaciones de 100, 200, 300 y 400 individuos, en la zona cafetera (Figura N°4.). El número de grupos resultó igual al número de preguntas. Se seleccionaron 250 observaciones de cada grupo haciéndose una matriz Z de $(250 \times 4 \times 4)$, el peso del exponente y el factor de determinación fue igual al del método anterior. El algoritmo Gustafson - Kessel se realizó utilizando también MATLAB 6.1. El resultado del algoritmo, es decir, los centros se resumen en la Figura N° 4 y las funciones de pertenencia en la Figura N°5. Al resolver la función objetivo no se observó ninguna pérdida en la clasificación.

FIGURA N°5
FUNCIÓN DE PERTENENCIA DETERMINADA POR EL ALGORITMO GUSTAFSON-KESSEL



Fuente: Elaboración propia.

II.4. Determinación cuantitativa de las cuatro preguntas.

Como se dijo anteriormente, el algoritmo FCM y el G-K fueron utilizados para cuantificar y clasificar en diferentes grupos información asimétrica. Basándose en estos resultados, se puede determinar información asimétrica en cada una de las preguntas formuladas. Como se muestra en la ecuación (9) (matriz de co-varianza de los grupos) la señal obtenida a partir de la encuesta es directamente proporcional al grupo de preguntas formuladas. Por ello, utilizando el prototipo o centro entregado por el GK o por FCM para 250 personas examinadas, es decir, luego de obtener la mínima función objetivo, $v_{i,250}$ la información asimétrica puede ser estimada como:

$$q_i = \frac{|z_i|}{\min |v_{i,250}|} \cdot 250 \tag{20}$$

Donde q_i indica el grado de desviación para el grupo de pacientes observados.

$$q_j = \frac{|z_i|}{\max |v_{i,250}|} \cdot 250 \tag{21}$$

Y q_j indica el grado de desviación para el grupo de médicos observados. Obteniendo así: los datos que muestra la Tabla N°3.

TABLA N°3
GRADO DE DESVIACIÓN O ASIMETRÍA DEL GRUPO

Parámetro del seguro	q_i		q_j	
Grado de desviación o asimetría del cluster	4.0910	1.8045	4.2484	79.3625

Fuente: Elaboración propia.

Para demostrar la aplicabilidad de este método, se determinó la asimetría de la información en la relación paciente médico. Esto se logró utilizando las ecuaciones (20) y (21). El resultado se resume en la Tabla N°3. La desviación estándar de z_i (que representa la información obtenida por la encuesta, es decir, información con contenido asimétrico), es igual a 57.8792, que es bastante similar al valor promedio extremo $(q_i + q_j)/2$, que es igual a 40.58. La información asimétrica es entonces, aproximadamente idéntica. Los valores característicos para salud, estatus social y demografía fueron clasificados bastante bien en datos provenientes del eje cafetero. Se observó una mezcla en la encuesta en cuanto a la decisión del paciente, fundamentada en el estado de salud, demografía y estatus socio económico. El método se aplicó dentro de los seis estratos socioeconómicos, pero solamente se resolvió el problema a través del método difuso. A pesar de que los valores propios en los grupos evaluados fueron diferentes, el estatus de salud fue bien clasificado y seleccionado; así mismo, se identificaron factores como pago de seguro y demografía.

La característica de los valores propios de cada grupo con diferente estructura y posición en la zona Colombiana seleccionada, tales como edad y sexo, fueron bien separados. Estas preguntas también fueron identificadas por el algoritmo FCM. El algoritmo GK probó ser bueno en las preguntas con similares estructuras tales como salud y seguro. Se diferenciaron cuatro subconjuntos a través del método de agrupamiento difuso.

III. CONCLUSIÓN

El objetivo que perseguimos con el presente trabajo es dar tratamiento a uno de los temas que es considerado de suma importancia en la administración: esto es, la forma de reducir la incertidumbre en la toma de decisiones empleando un método diferente al tradicional.

Para dicho tratamiento se emplearon algunas herramientas estadísticas y otras de la lógica borrosa. Se pretendió dejar en evidencia la importancia de los dos algoritmos aquí tratados para el estudio de temas de administración, así también se mostró la forma en que se complementan.

La aplicación del método de agrupamiento borroso tiene grandes ventajas en la determinación de factores que influyen en la decisión del paciente. Propiedades tales como demografía, estrato socioeconómico, estado de salud, etc. influyen en dicha

decisión. Sin embargo, no puede ser exactamente conectado a la señal que entrega una encuesta, en un proceso de muestreo aleatorio simple. Se inspeccionó el tipo de respuesta y se analizó bajo estas condiciones. Se incorporó conocimiento a priori observando mejoras en el algoritmo de agrupamiento (en la aplicación de los algoritmos FCM y GK). Esto prueba ser una vía interesante para la clasificación y cuantificación de la asimetría y de los factores que determinan la decisión de un paciente para visitar al médico.

Como puede observarse en la Gráfica N°3, la geometría que esboza el grupo en relación al centro puede estar o no deformada. Estas asimetrías se pueden asociar con la asimetría de la información. Esto se logró con la expresión presentada en (20 y 21) y haciendo un promedio de la valoración de los q_i contra los q_j .

El resultado fue expresado en la Tabla 2: los actores en el marco del modelo "principal-agente", muestran claramente teoría de información asimétrica y es evidente que existe una variable no verificable en este contrato y que afecta en sentido contrario los intereses de los participantes en esta relación. Esto se observa en los renglones impares de la Tabla 2 que corresponden a los agentes. Los renglones pares corresponden a los principales, en la misma Tabla 2.

Medir esta asimetría se logró utilizando las expresiones representadas en las ecuaciones (20) y (21). La Tabla 3 mostró el resultado al aplicar la evaluación de asimetría.

La variable I es un elemento altamente influenciado por perturbaciones exógenas y variables ocultas resultado de la politización. Esta situación queda detallada por el bajo porcentaje indicado en la columna de la Tabla N° 3, concerniente a este parámetro. Aunque la metodología y el origen de datos en la tesis de Vera, puede permitir valorar este resultado para España, pues osciló entre el 4% y el 16.5% como lo acota él en la Tabla 3 (Duda, Peter & David G, 1989). Para la evaluación realizada en el presente trabajo, vemos que los centros según G K es de 1.8% y según FCM observó una variación 1.77% al 2.34%. Esto es una evidencia de la diferencia entre la cultura del seguro que se vive en Colombia en relación a la española.

Este documento propone un relación geométrica proveniente del algoritmo de clasificación borroso, con el objeto de detallar los factores influyentes en la aplicación de una encuesta en relación de la visita del paciente al médico (se tuvo en cuenta también la calidad de paciente del médico), dando cumplimiento de esta manera al objetivo propuesto.

La finalidad no fue hacer un análisis exhaustivo del tema, tarea que se deja para otra oportunidad o para otros investigadores, sino que se prefirió dar un panorama a manera de introducción de la forma en que puede ser analizado el problema asimétrico. Es importante centrarse en la utilidad que tiene la lógica borrosa en estos casos.

REFERENCIAS

- DEL BRIO, B. & A. SANZ. 2007. Redes Neuronales y Sistemas Borrosos, Ed. Alfaomega.
- DÍEZ, J. & NAVARRO, J. 2004. Algoritmos de Agrupamiento en la identificación de modelos Borrosos . Sala Departamento de Ingeniería de Sistemas y Automática Universidad Politécnica de Valencia, CEA-IFAC

- DUDA, R.; P. HART & D. STORK. Fuzzy control and fuzzy system, Witold Pedrycz. Ed. Research Studie Press Ltd.
- EDWARDS, R.E. 1995. Funcional Analysis: theory and applications. Dover Publications, Inc. New York.
- GIL ALUJA, J. 2002. Introducción de la teoría de la incertidumbre en la gestión de empresas. Ed. Milladoiro-Academia de Doctors. Vigo-Barcelona (España).
- GILES, D. & R. DRAESEKE. 2001. Econometric Modelling Based on Pattern Recognition via the Fuzzy c-Means Clustering Algorithm, Department of Economics University of Victoria, Working Paper.
- GULLEY, N. JANG, J. 2000. Fuzzy Logic Toolbox User's Guide, The Math Works Inc. Massachusetts.
- KIM, Y.; D-W. KIM; D. LEE & K. LEE. 2004. A cluster validation index for GK cluster analysis based on relative degree of sharing q . Information Sciences. vol. 168 pp. 225-242. Elsevier Science Ltd
- KOBAYASHI, T. & T. AISHIMA. 1995. Relating sensory properties of tea aroma to gas chromatographic data by chemometric calibration methods. N. Food Research international, Vol. 28, No. 5, pp. 485-493. Elsevier Science Ltd.
- LIN, C-T. & G. LEE. 1995. Neural Fuzzy Systems: a neuro-fuzzy synergism to intelligent systems. Ed. Prentice Hall.
- LIN. H-C; F-C. LIN; T-Y. HSIAO; Y-C. LIN. 2008. Fuzzy set theory in managerial contract analyses: Expert Systems with Applications.
- LOZANO. M. & F. FUENTES. 2003. Tratamiento borroso del intangible en la valoración de empresas de internet. <http://eumed.net/cursecon/librería/index.htm>.
- MATHWORKS. 2000. MATLAB 6.1.
- SHAO, Q. & Y. WU. 2005. A consistent procedure for determining the number of clusters in regression clustering. Journal of Statistical Planning and Inference vol. 135 pp. 461 - 476. Elsevier Science Ltd.
- VERA, A. 2001. Microeconometrics and Asymmetric Information: aplicaciones to health care utilization. Dissertation in economics: Universitat Autònoma de Barcelona Belaterra, March.