

Modelado Causal en Marketing mediante Aprendizaje no Supervisado de Reglas de Asociación Difusas

Albert Orriols-Puig¹ Jorge Casillas² Francisco J. Martínez-López³

¹ Grup de Recerca en Sistemes Intel·ligents, Enginyeria i Arquitectura La Salle,
Universitat Ramon Llull, 08022 Barcelona, aorriols@salle.url.edu

² Dpto. Ciencias de la Computación e Inteligencia Artificial,
Universidad de Granada, 18071 Granada, casillas@decsai.ugr.es

³ Dpto. Comercialización e Investigación de Mercados,
Universidad de Granada, 18071 Granada, fjmlopez@ugr.es

Resumen

En marketing es habitual que se haga uso de modelos complejos, con múltiples relaciones de dependencia, para aproximarse al entendimiento diversos fenómenos de su interés. Este tipo de modelos son especialmente utilizados en la disciplina del comportamiento del consumidor. Están formados por lo que se conoce como variables latentes o constructos, esto es, variables que tienen un nivel de abstracción tal que no permiten medirse de manera unívoca por una única evaluación del consumidor, sino que precisan ser medidas por varias evaluaciones a través de escalas multi-ítem. Este hecho añade incertidumbre a las variables. Recientemente, algunos autores han planteado soluciones metodológicas basadas e inteligencia artificial que permiten trabajar con este tipo de dato, permitiendo, por tanto, aplicarse en el modelado de marketing. El trabajo presente sigue esta línea y propone afrontar el problema con Fuzzy-CSar, un sistema clasificador que extrae reglas de asociación difusas para el modelado del comportamiento del consumidor. La novedad de nuestra aproximación con respecto a técnicas existentes es que no se presupone ningún modelo causal previo, sino que se buscan relaciones fuertes entre las variables del sistema. Fuzzy-CSar se aplica a un problema de marketing real y los resultados se comparan con los obtenidos con un sistema GCCL diseñado expresamente para el mismo problema. Los resultados muestran las ventajas de evolucionar reglas de asociación y la competitividad de Fuzzy-CSar en general.

Palabras Clave: Lógica difusa; algoritmos genéticos; reglas de asociación; Marketing; comportamiento del consumidor

1 INTRODUCCIÓN

En las disciplinas de empresa y, en particular, en la de marketing, el uso de modelos ha sido tradicional para dirigir los procesos de análisis de datos. Como sabemos, este hecho supone que una estructura de relación entre los elementos del modelo (es decir, variables), conocida a priori, se utilice para, por medio de métodos analíticos de datos, describir o predecir la relación entre los mismos. Este enfoque de análisis de los datos respeta el procedimiento clásicamente establecido por el método científico, por el que un investigador previamente plantea unas hipótesis de relación entre variables para, posteriormente, utilizar métodos de análisis mediante los que contrastarlas en los datos y, finalmente, extraer conclusiones de los resultados obtenidos. Básicamente, en eso consiste el proceso de modelado en marketing, siendo habitual su utilización en el ámbito práctico/profesional por parte de los analistas de mercados. Esto es, se utiliza un modelo para dirigir el proceso de búsqueda de información en las bases de datos con el objeto de apoyar decisiones de marketing.

Pero, dicho lo anterior, ¿podría prescindirse de los modelos? Si bien la utilidad de los modelos es indudable, paradójicamente, su uso puede limitar el proceso de agregación de valor a los datos en ciertos tipos de problemas de decisión. Esto tiene implicaciones sobre todo en el ámbito práctico, en el de la toma de decisiones en marketing, donde los procesos de búsqueda de información están mucho menos condicionados por los procedimientos metodológicos, si lo comparamos con el académico, y mucho más por la calidad de la información generada. En otras palabras, en determinadas circunstancias, el conocimiento que generan métodos analíticos basados en estructuras relacionales de variables puede no tener toda la efectividad que requeriría un gestor de marketing. Esto se debería a que la estructura de base utilizada para dirigir el proceso de extracción de conocimiento de los datos y, conse-

cuentemente, la estructura de la información generada, no es adecuada para resolver el problema de decisión.

En particular, existen determinados escenarios de decisión, caracterizados por estar regular o, en su caso, deficientemente estructurados o delimitados, donde no existe información a priori de base para orientar el proceso de búsqueda de información, etc., en los que se precisa un cambio de mentalidad a la hora de desarrollar y aplicar los métodos analíticos de bases de datos. Para apoyar la toma de decisiones en marketing se utilizan multitud de métodos, importados generalmente de otras disciplinas, de entre las que destacan la estadística y la inteligencia artificial. En cierta medida, la evolución de los sistemas de apoyo a la gestión de marketing (SAGMk) se comprende por el tipo de métodos en los que se basaran principalmente. No obstante, desde finales de los 80, estos sistemas han mostrado una tendencia clara hacia lo que se denominan SAGMk basados en el conocimiento, gran categoría de agrupación de sistemas que se caracterizan por utilizar métodos analíticos basados en inteligencia artificial [1]. Sin duda, estamos convencidos de que el presente y futuro de las soluciones para este tipo de problemas de decisión que comentamos se encuentra en esta disciplina, concretamente en el campo de la extracción de conocimiento de bases de datos (KDD). En este sentido, el proceso de KDD y, en particular, en su etapa de minería de datos, se pueden distinguir dos grandes estilos de ejecución, así como un tercero fruto de la combinación de los mismos [7]: el enfoque “top-down”, donde el experto parte de una estructura relacional a priori entre las variables que conforman la base de datos a analizar; y el enfoque “bottom-up”, donde, en contraposición al anterior, no se hace uso de ninguna información a priori sobre la estructura, utilizándose métodos que exploran automáticamente los datos y que sugieren los patrones más significativos subyacentes a los mismos. Es lógica, por tanto, la asociación del primer y el segundo enfoque con los estilos de aplicación en minería de datos supervisada y no supervisada respectivamente.

Puede que los métodos de KDD basados en un enfoque “bottom-up” parezcan poco ortodoxos, e incluso poco útiles a primera vista. En cierta medida, se puede pensar que las soluciones de este tipo de métodos implicarían que habría que encontrar sentido a muchos patrones de información obtenidos, pudiendo resultar algo caótico. No obstante, no perdamos de vista el tipo de problemas de decisión donde sería más procedente su uso: los problemas poco o mal estructurados, donde es más difícil orientar el sentido de las decisiones, puesto que el conocimiento previo del que se dispone no permite abordar su resolución de manera inmediata. Bajo este escenario, los patrones extraídos de la

base pueden ayudar precisamente a encontrar sentido al problema de información que se aborda. En cierta medida, los resultados de métodos de KDD diseñados con este enfoque actuarían estimulando la creatividad del usuario/tomador de decisiones de marketing que se apoye en ellos. Por tanto, utilizando la clasificación ORAC propuesta por Wierenga y Van Bruggen [5, 6] para la resolución de problemas en marketing, este tipo de soluciones podrían encuadrarse bajo el modo creativo. Este modo de resolución de problemas se caracteriza por propiciar un razonamiento divergente, esto es, abierto a todas las opciones posibles, contemplando incluso otras soluciones aparentemente no válidas o descabelladas si se aplicara un razonamiento, más ortodoxo, convergente.

En este trabajo proponemos un sistema de aprendizaje evolutivo, con una aproximación descriptiva, para bases de datos de marketing; en particular, la aplicación la realizamos sobre una base de datos que recoge variables de comportamiento del consumidor. En concreto, el diseño del algoritmo se ha realizado para un escenario de decisión caracterizado por los siguientes puntos: (1) se parte de un conjunto de variables cuya relación estructural se desconoce o se percibe deficiente, por lo que la utilización de un modelo para la extracción de conocimiento no es posible o, en su caso, conveniente; (2) el gestor de marketing está interesado especialmente en una variable objetivo, que quiere explicar en función de las anteriores, aunque la extracción de reglas que expliquen el comportamiento de otras variables también se considera valioso. Para cumplir este objetivo, proponemos Fuzzy-CSar un sistema clasificador basado en algoritmos genéticos que evoluciona un conjunto de reglas de asociación. El sistema usa una representación difusa que le permite tratar con datos imprecisos, los cuales caracterizan el problema de marketing abordado. De esta forma, cada variable latente se compone de un conjunto de variables observadas (ítemes) que aportan información parcial sobre la misma. Por tanto, la información de cada ítem debe ser agregada para definir la variable en cuestión. Para este fin, se usa el emparejamiento multi-ítem definido en [3]. Los resultados experimentales y la comparación del sistema con una solución GCCL [3] expresamente diseñada para encontrar un conjunto de reglas que definan una sola variable objetivo demuestran la robustez de nuestra propuesta.

La organización del resto del artículo es la siguiente. La sección 2 presenta brevemente los conceptos básicos de reglas de asociación. La sección 3 describe Fuzzy-CSar en detalle y la sección 4 aplica el sistema para extraer reglas de asociaciones interesantes del problema de marketing. Finalmente, la sección 5 concluye

el trabajo y presenta líneas de trabajo futuro.

2 REGLAS DE ASOCIACIÓN DIFUSAS

Dado un conjunto de ejemplos $I = \{i_1, i_2, \dots, i_n\}$, un conjunto de atributos $A = \{a_1, a_2, \dots, a_\ell\}$ y una colección de conjuntos difusos asociados a cada variable $F_{a_i} = \{f_{i1}, f_{i2}, \dots, f_{im}\}$, una regla de asociación difusa presenta la siguiente estructura [2]:

$$\begin{aligned} \text{si } X = \{x_1, x_2, \dots, x_p\} \text{ es } \{F_{x_1}, F_{x_2}, \dots, F_{x_p}\} \\ \text{entonces } Y = \{y_1, y_2, \dots, y_p\} \text{ es } \{G_{x_1}, G_{x_2}, \dots, G_{x_p}\}, \end{aligned} \quad (1)$$

donde X e Y son dos conjuntos de atributos disjuntos tal que $X \in A$ e $Y \in A$.

El objetivo de los sistemas de aprendizaje de reglas de asociación es generar reglas que denoten asociaciones interesantes entre las variables del problema. Aunque se pueden encontrar muchas medidas de interés en la literatura, las dos más usadas son el soporte (*sop*), un estimador de la frecuencia de aparición del concepto descrito en la regla, y la confianza (*conf*), un estimador de la veracidad de la asociación descrita. En este trabajo, las dos métricas se calculan como

$$\begin{aligned} \text{sop}(\mathbf{R}) &= \frac{1}{N} \sum_{e=1}^N \mu_A(x^{(e)}) \cdot \mu_B(y^{(e)}) \quad (2) \\ \text{conf}(\mathbf{R}) &= \frac{\sum_{e=1}^N (\mu_A(x^{(e)}) \cdot \max\{1 - \mu_A(x^{(e)}), \mu_B(y^{(e)})\})}{\sum_{e=1}^N \mu_A(x^{(e)})} \quad (3) \end{aligned}$$

donde μ_A y μ_B es el grado de emparejamiento de la regla con el ejemplo correspondiente (ver la siguiente sección para detalles de los cálculos).

3 DESCRIPCIÓN DE FUZZY-CSAR

Fuzzy-CSar es una arquitectura incremental de evaluación de reglas de asociación difusas que usa un algoritmo genético (AG) para el descubrimiento de reglas interesantes, es decir, reglas con alto soporte y confianza. A continuación, se describen en detalle los diferentes componentes del sistema.

3.1 REPRESENTACIÓN DEL CONOCIMIENTO

Fuzzy-CSar evoluciona un conjunto o *población* de individuos o *clasificadores*, los cuales están formados por una regla de asociación difusa ($X \rightarrow Y$) y una serie de parámetros. La regla difusa se representa del siguiente modo:

$$\text{si } x_i \text{ es } \widetilde{A}_i^k \text{ y } \dots \text{ y } x_j \text{ es } \widetilde{A}_j^k \text{ entonces } y_c \text{ es } \widetilde{B}_c^k,$$

donde el antecedente X es representado por un conjunto de ℓ_a variables $x_i \dots x_j$ ($0 < \ell_a < \ell$, siendo

ℓ el número de variables del problema) y el consecuente Y contiene una única variable y_c que no existe en el antecedente. Por consiguiente, se permite que las reglas tengan un número arbitrario de variables en el antecedente. Cada variable se representa con una disyunción de etiquetas lingüísticas $\widetilde{A}_i^k = \{A_{i1} \wedge \dots \wedge A_{in_i}\}$. Para evitar tener reglas con variables muy generales que aporten poca información al experto humano, el sistema permite fijar el número máximo de etiquetas lingüísticas permitidas en las variables del antecedente (*maxEtiqAnt*) y del consecuente (*maxEtiqCons*).

Cada clasificador tiene asociado seis parámetros: (1) el soporte *sop*, un indicador de la frecuencia de aparición de la regla y de su grado de emparejamiento con los ejemplos de entrada; (2) la confianza *conf*, que indica la fuerza de la implicación; (3) el *fitness* F , que denota la cualidad de dicha regla; (4) la experiencia *exp*, el número de veces que la regla se ha emparejado con grado mayor que cero con un ejemplo de entrada; (5) la numerosidad *num*, el número de copias del clasificador en la población; y (6) el tamaño del conjunto de asociación *as*, la media de los tamaños de los conjuntos de asociación en los que la regla ha participado.

3.2 EMPAREJAMIENTO MULTI-ÍTEM

Fuzzy-CSar usa el concepto de *emparejamiento multi-ítem* propuesto en [3] para considerar apropiadamente el conjunto de ítems que representan cada variable. La idea es que se une la información aportada para cada ítem de una misma variable. El método computa el grado de emparejamiento de una variable i con el atributo de entrada correspondiente e_i como

$$\mu_{\widetilde{A}_i}^{\sim}(e_i) = \max_{h_i=1}^{p_i} \mu_{\widetilde{A}_i}^{\sim}(x_i^{(e_i)}), \quad (4)$$

donde p_i es el número de ítems que representan la variable i . Así pues, el grado de emparejamiento de la variable es el grado máximo de emparejamiento de sus ítems.

Además, el emparejamiento multi-ítem trata especialmente las variables/constructos de segundo orden, esto es, variables que se reflejan en otros de primer orden, en lugar de en ítems. Debido a que las variables de segundo orden representan la intersección de distintas variables de primer orden, se calcula el grado de este tipo de variables como la T-norma (mínimo) de cada conjunción difusa, es decir,

$$\mu_{\widetilde{A}_i}^{\sim}(e_i) = \min_{k_i=1}^{s_i} \mu_{\widetilde{A}_i}^{\sim}(x_i^{(e_i)}). \quad (5)$$

3.3 INTERACCIÓN DE APRENDIZAJE

En cada iteración de aprendizaje, Fuzzy-CSar recibe un ejemplo de entrada $e = (e_1, e_2, \dots, e_\ell)$. Seguida-

mente, el sistema crea el *conjunto de emparejamiento* [M] con todos los clasificadores de la población que tienen un grado de emparejamiento con e superior a 0. Si [M] tiene menos de θ_{mna} clasificadores, se dispara el *operador de cubrimiento*, el cual genera nuevas reglas hasta que [M] tenga θ_{mna} clasificadores. A continuación, se crean distintos *conjuntos de asociación* [A] con los clasificadores de [M] siguiendo la metodología que se detalla en el apartado 3.3.2. Se da una probabilidad de selección a cada [A] proporcional a la media de la confianza de los clasificadores de dicho [A]. Luego, se aplica el operador de subsumisión al [A] seleccionado (ver apartado 3.3.3) y se actualizan los parámetros de los clasificadores en [P]. Al final de la iteración se aplica el AG sobre [A] si el tiempo medio desde su última aplicación sobre dicho [A] es superior a θ_{GA} . A continuación se explican más detalladamente (1) el operador de cubrimiento, (2) la metodología de creación de conjuntos de asociación, (3) la subsumisión en el conjunto de asociación, y (4) la actualización de parámetros.

3.3.1 Operador de Cubrimiento

El operador de cubrimiento usa el ejemplo de entrada e para generar un clasificador con el máximo grado de emparejamiento posible con e . Para cada variable, el operador decide aleatoriamente (con probabilidad $1 - P_{\#}$) si la variable debe estar en el antecedente de la regla, con la restricción de que se deben seleccionar un mínimo de 1 variable y un máximo de $\ell - 1$ variables. Seguidamente, se escoge una de las variables no seleccionadas para formar el consecuente de la regla. Cada una de las variables seleccionadas se inicializa con la etiqueta lingüística que maximiza el grado de emparejamiento con el valor correspondiente del atributo (e_i). Finalmente, las variables se generalizan introduciendo cada etiqueta lingüística no seleccionada con probabilidad $P_{\#}$.

3.3.2 Creación de los Conjuntos de Asociación

Este procedimiento persigue crear conjuntos de reglas con características similares para establecer una competición entre ellas con el objetivo que solamente se mantengan las reglas de más calidad en la población. Persiguiendo este fin, Fuzzy-CSar aplica el siguiente procedimiento heurístico. Primeramente, se hace una ordenación ascendente de las reglas de la población según la variable que tienen en el consecuente. Dadas dos reglas r_1 y r_2 que tienen la misma variable en el consecuente, se considera que $r_1 < r_2$ si $\ell_1 < \ell_2$ o ($\ell_1 = \ell_2$ y $u_1 > u_2$), donde ℓ_1 , u_1 , ℓ_2 y u_2 son, respectivamente, la posición del centro del primer y último conjunto difuso de las variables del consecuente de las reglas r_1 y r_2 .

A continuación se construyen los conjuntos de asociación. Se crea el primer [A] y se le inserta el primer clasificador de la población. Entonces, el siguiente clasificador k también se inserta si tiene la misma variable en el consecuente y ℓ_k es inferior que el u_i mínimo entre todos los clasificadores de [A]. El proceso se repite hasta que se encuentra el primer clasificador que no satisface la condición. En este caso, se crea un [A] nuevo y se aplica el mismo proceso para insertar clasificadores en él. La idea que reside en esta estrategia es que las reglas que explican la misma región de un mismo consecuente seguramente denotan las mismas asociaciones entre las variables de entrada.

3.3.3 Subsumisión en el Conjunto de Asociación

Fuzzy-CSar aplica un proceso de subsumisión de reglas con el objetivo de eliminar reglas muy parecidas de la población. Para cada regla del conjunto de asociación seleccionado, se comprueba que la regla no pueda ser subsumida por ninguna de las otras reglas que están en el mismo conjunto de asociación. Se considera que una regla r_1 subsume a otra regla r_2 si se cumplen las cuatro condiciones siguientes: (1) r_1 tiene suficiente experiencia y alta confianza ($conf_1 > conf_0$ y $exp_1 > \theta_{exp}$, donde $conf_0$ y θ_{exp} son dos parámetros de configuración); (2) todas las variables presentes en el antecedente de r_1 también existen en el antecedente de r_2 (además, r_2 puede tener otras variables en el antecedente); (3) la variable en el consecuente de r_1 y r_2 es la misma; y (4) r_1 es más general que r_2 , es decir, para cada variable en el antecedente o consecuente de r_1 , esta variable tiene definidas, como mínimo, las mismas etiquetas lingüísticas que la misma variable en r_2 .

3.3.4 Actualización de Parámetros

Al final de cada iteración de aprendizaje se actualizan los parámetros de todos los clasificadores de la población. Primeramente, se actualiza el soporte de la regla según

$$sop_{t+1} = \frac{sop_t \cdot (lt - 1) + \mu_{\tilde{A}}(x^{(e)}) \cdot \mu_{\tilde{B}}(y^{(e)})}{lt}, \quad (6)$$

donde lt es el tiempo de vida del clasificador, es decir, el número de iteraciones que ha estado en la población. A continuación, se actualiza la confianza de la regla:

$$conf_{t+1} = \frac{sum_imp_{t+1}}{sum_mat_{t+1}}. \quad (7)$$

sum_imp y sum_mat se definen como

$$sum_imp_{t+1} = sum_imp_t + \mu_{\tilde{A}}(x^{(e)}) \cdot \max\{1 - \mu_{\tilde{A}}(x^{(e)}), \tilde{B}(y^{(e)})\} \quad (8)$$

$$sum_mat_{t+1} = sum_mat_t + \mu_{\tilde{\lambda}}(x^{(e)}). \quad (9)$$

Seguidamente, se computa el *fitness* de cada clasificador:

$$F = conf^{\nu}. \quad (10)$$

Finalmente, se actualiza *as* el cual mantiene la media aritmética del número de clasificadores de los conjuntos de asociación en los que el clasificador ha participado.

3.4 DESCUBRIMIENTO DE REGLAS

Fuzzy-CSar usa un AG *steady-state* para crear nuevas reglas prometedoras. El AG selecciona dos clasificadores con probabilidad proporcional a su *fitness*. Estos dos padres se cruzan para generar dos hijos cl_1 y cl_2 con probabilidad P_{chi} ; de lo contrario, los hijos son una copia exacta de los padres. Fuzzy-CSar utiliza un cruce uniforme que contempla la restricción que los hijos deben tener, como mínimo, una variable en antecedente. Los hijos resultantes pueden sufrir tres tipos distintos de mutación: (1) introducción/borrado de variables en antecedente (con probabilidad $P_{I/B}$); (2) mutación de las etiquetas lingüísticas de variables existentes (con probabilidad P_{μ}); (3) mutación de la variable de consecuente (con probabilidad P_C). El primer tipo de mutación selecciona aleatoriamente si se debe añadir una de las variables no usadas (sólo aplicable si hay variables no usadas en la regla) o quitar una variable del antecedente (sólo aplicable si el antecedente tiene más de una variable). El segundo tipo de mutación modifica los términos lingüísticos de una de las variables de la regla acorde con el método definido en [3]. El tercer tipo de mutación cambia la variable del consecuente, escogiendo aleatoriamente una de las posibles variables de salida. En cualquiera de los tres tipos de mutación, cuando se añade una nueva variable, esta se inicializa con la etiqueta lingüística que maximiza el grado de emparejamiento con el atributo correspondiente del ejemplo de entrada.

Después del cruce y de la mutación, los descendientes se introducen en la población, borrando clasificadores si se sobrepasa su tamaño máximo. En este caso, los clasificadores reciben una probabilidad de borrado proporcional a su parámetro *as*. Además, esta probabilidad de borrado se aumenta si el *fitness* del clasificador es inferior a $\delta \bar{F}$, donde \bar{F} es el *fitness* medio de la población y δ es un parámetro de configuración.

3.5 REDUCCIÓN DE LA POBLACIÓN

Al final del proceso de aprendizaje se usa la siguiente heurística para eliminar reglas poco interesantes de la

población final. En primer lugar, se eliminan todas las reglas cuya experiencia es inferior a θ_{exp} . Entonces, para cada regla en la población, se mira si existe otra regla que la pueda subsumir. Se usa el mismo criterio que el mencionado en el apartado 3.3.3 para decidir si una regla r_1 puede subsumir otra regla r_2 pero con una excepción: se requiere que la confianza de r_1 sea mayor que la confianza de r_2 en vez de requerir que la confianza de r_1 sea mayor que $conf_0$.

4 EXPERIMENTACIÓN

4.1 DESCRIPCIÓN DEL PROBLEMA

La experimentación parte del modelo causal presentado en [4], el cual se basa en el análisis del estado de *flow* del usuario al navegar por la web. Se han utilizado los datos proporcionados por los mismos autores; esta práctica es habitual en escenarios de investigación en los que se evalúa empíricamente el rendimiento de una herramienta analítica basada en inteligencia artificial para modelado en marketing. El concepto de *flow* en Internet se alcanza cuando el usuario esta totalmente inmerso en la navegación, hasta tal punto que “nada más parece importarle” [4].

El modelo que consideramos en la experimentación consta de 9 variables: *inicio de uso de la web*, *implicación en la web*, *habilidad/control*, *velocidad de interacción*, *desafío/estímulo*, *nivel de atención*, *telepresencia/distorsión de tiempo* y *flow*. Todas las variables toman valores enteros de 1 a 9 excepto *inicio de uso de la web*, que puede tomar intervalos de años de experiencia como usuario de la Web, cuyas marcas de clase con las siguientes: {0,25, 0,75, 1,50, 2,50, 3,50 y 4,50}. Se ha usado una semántica difusa con tres etiquetas lingüísticas triangulares distribuidas uniformemente para todas las variables del sistema excepto *inicio de uso de la web*. Para esta última variable se usa una semántica no uniforme con seis etiquetas lingüísticas, cada una centrada en uno de los seis posibles valores de la variable.

Se han hecho tres tipos distintos de experimentos con este problema fijando el grupo de variables que podían estar en el antecedente y el grupo de variables que podían estar en el consecuente.

- En el primer experimento (*4in1out*), se han considerado únicamente cuatro de las nueve posibles variables en el antecedente (*habilidad/control*, *velocidad de interacción*, *desafío/estímulo* y *telepresencia/distorsión de tiempo*) pues son las variables que tienen una relación directa con *flow* según el modelo causal proporcionado en [4].
- En el segundo experimento (*9in1out*), se ha con-

siderado que las reglas de asociación podían contener todas las variables en el antecedente excepto *flow* y el consecuente se restringe a la variable *flow*.

- En el tercer experimento (*9in9out*), se permite que cualquiera de las nueve variables (incluyendo *flow*) puedan estar tanto en el antecedente como en el consecuente de la regla.

Los resultados obtenidos con Fuzzy-CSar se han comparado con los proporcionados por el sistema GCCL presentado en [3], un sistema multi-objetivo basado en NSGA-II que obtiene un frente de Pareto con las mejores reglas encontradas (esto es, las reglas no dominadas en términos de soporte y confianza). Es importante destacar la diferencia entre ambas aproximaciones: mientras la solución GCCL tiene como objetivo optimizar el frente Pareto, el objetivo de Fuzzy-CSar es evolucionar un conjunto de reglas de asociación diversas que tengan máxima confianza. Los dos sistemas se han configurado como se especifica a continuación. Para Fuzzy-CSar se han usado los siguientes parámetros: tamaño_pop = 6.400, $\theta_{GA} = 50$, $P_{\#} = 0,5$, $\theta_{exp} = 1.000$, $conf_0 = 0,95$, $\nu=1$, $P_{\chi} = 0,8$, $P_{I/B} = 0,1$, $P_{\mu} = 0,1$, $P_C = 0,1$, $\delta = 0,1$. Para el sistema GCCL, se ha usado (ver [3] para detalles sobre la notación): número de generaciones 100, tamaño de la población 100, probabilidad de cruce 0,7 y probabilidad de mutación por cromosoma 0,1. Todos los resultados que se presentan a continuación son medias de 10 semillas aleatorias distintas.

4.2 RESULTADOS

La figura 1 compara los frentes Pareto (media de 10 ejecuciones) obtenidos con los sistemas Fuzzy-CSar y GCCL cuando se usan ocho variables en el antecedente y la variable *flow* en el consecuente (problema *8in1out*). Los resultados son equivalentes a usar 4 variables en el antecedente (problema *4in1out*); por falta de espacio y por similitud con la figura 1, se omite la figura de los frentes Pareto para el problema *4in1out*. Esta similitud entre los frentes Pareto generados con ambos problemas soporta la hipótesis que las cuatro variables escogidas en el primer experimento son las más relevantes para definir el concepto de *flow*.

La tabla 1 complementa los resultados mostrando el tamaño de la población evolucionada por Fuzzy-CSar, el número de reglas no dominadas dentro de esta población y el tamaño del conjunto Pareto creado por GCCL; asimismo, la tabla 2 indica la distancia de *crowding* media entre las soluciones de la población para cada caso. Los resultados muestran claramente que Fuzzy-CSar evolucionar un frente Pareto muy similar al obtenido por GCCL; en ambos casos, las solu-

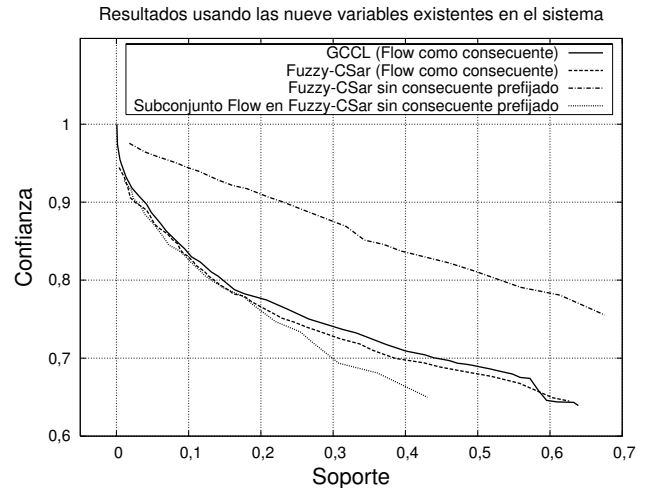


Figura 1: Comparación del frente Pareto medio de Fuzzy-CSar y GCCL

Tabla 1: Número medio de reglas evolucionadas por Fuzzy-CSar, número de estas reglas que están en el Pareto y tamaño del Pareto obtenido por GCCL

Problema	FCSar	FCSar Par.	GCCL Par.
4in1out	479,2	76,3	82,6
8in1out	1259,7	105,9	84,4
9in9out	1752,5	468,3	—

Tabla 2: Media de la distancia de *crowding* de los frentes Pareto evolucionados con Fuzzy-CSar y GCL

Problema	FCSar	FCSar Par.	GCCL Par.
4in1out	$2.36 \cdot 10^{-3}$	$1.53 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$
8in1out	$9.54 \cdot 10^{-4}$	$1.07 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$
9in9out	$6.88 \cdot 10^{-4}$	$2.58 \cdot 10^{-3}$	—

ciones están uniformemente distribuidas a lo largo del Pareto. Además, Fuzzy-CSar también genera un conjunto variado de reglas (479 y 1260 en media para los problemas *4in1out* y *9in1out* respectivamente) que, aun no estando en el Pareto, pueden poseer conocimiento interesante para el experto humano. Cabe destacar las diferencias entre ambos sistemas para ver el significado de estos resultados. Fuzzy-CSar es un sistema incremental de aprendizaje que utiliza un sistema implícito de *niching* genotípico y optimiza en función de la confianza de cada regla; así pues, Fuzzy-CSar no tiene el objetivo implícito de optimizar el frente Pareto, sino de mantener una solución distribuida en el espacio de búsqueda en base a nichos. En cambio, el sistema GCCL es un multi-objetivo que persigue la optimización del frente Pareto en base dos objetivos: confianza y soporte. Para conseguir este fin, GCCL usa ordenación por *ranking*, *niching* explícito

mediante una función de *crowding* y un esquema de selección de torneo basado en la ordenación de las soluciones en distintos frentes. Por tanto, el hecho que Fuzzy-CSar presente un frente Pareto similar a un sistema especializado en optimizar el frente Pareto como es GCCL confirma la robustez Fuzzy-CSar.

En segundo lugar, la figura 1 también recoge los resultados del tercer experimento, esto es, cuando se permite que cualquier variable esté en el antecedente o consecuente de las reglas de Fuzzy-CSar; en este caso, Fuzzy-CSar se usa para buscar asociaciones entre cualquier variable del problema. Por consiguiente, en esta aproximación dejamos que Fuzzy-CSar descubra las asociaciones más importantes, esto es, el modelo estructural del sistema, sin considerar el modelo causal proporcionado inicialmente. Los resultados muestran el potencial diferenciador de nuestra aproximación. En una misma ejecución, Fuzzy-CSar es capaz de evolucionar un conjunto de reglas con alto soporte y confianza, las cuales tienen distintos consecuentes, resultando en un frente Pareto que supera ampliamente al obtenido con el método GCCL. Por último, la figura muestra el frente Pareto de esta ejecución cuando sólo se consideran las reglas que tienen la variable *flow* en el consecuente. En este caso, el Pareto se asemeja al original por valores de confianza altos y se ve ligeramente degradado a medida que la confianza de las soluciones decrece. Este comportamiento se puede explicar con la siguiente hipótesis. En este experimento, el número de reglas que Fuzzy-CSar debe evolucionar crece con el número de consecuentes considerados, lo que promueve la competición entre reglas para mantenerse en la población. Como Fuzzy-CSar optimiza en función de la confianza de las reglas, el sistema prioriza las reglas con máxima confianza, lo que va en detrimento de reglas con menos confianza pero con más soporte.

Se podrían obtener resultados similares con la aproximación GCCL mediante la ejecución de nueve experimentos, considerando cada vez una de las nueve variables como salida. Esto resultaría en nueve Paretos distintos que se deberían juntar para obtener el Pareto final. Aún así, es importante destacar que Fuzzy-CSar proporciona un soporte natural para la extracción automática de reglas de asociación interesantes con distintos consecuentes, evolucionando paralelamente las distintas soluciones distribuidas en el espacio de características y manteniendo aquellas que tienen confianza más alta.

4.3 EJEMPLO DE INTERPRETACIÓN DEL CONOCIMIENTO EXTRAÍDO

Finalmente, de las múltiples reglas generadas en el proceso de minería de datos, seleccionamos una muestra ilustrativa del atractivo de la información ofrecida por

el método de aprendizaje no supervisado que presentamos. En concreto, sin perjuicio del interés de otras reglas generadas en las que encontramos una estructura de antecedentes y consecuentes más acorde con el modelo estructural finalmente propuesto en [4], a continuación mostramos algunos casos con una relación de antecedentes y consecuente no contemplada en dicho modelo. Precisamente, éste es el atractivo de las salidas ofrecidas por nuestro método, al no condicionarse el proceso de aprendizaje automático por ninguna estructura de conocimiento *a priori*.

R₁: SI implicaciónWeb es Media y habilidad/control es {Pequeña o Media} y nivelAtención es {Pequeño o Medio} y flow es {Pequeño o Medio} ENTONCES comportamExploratorio es Medio [Sop: 0,22; Conf: 0,87]

En primer lugar, la regla R₁ muestra una relación de antecedentes para el consecuente “comportamiento exploratorio”. En el modelo propuesto en [4] sólo existía un determinante directo de este tipo de comportamientos de navegación del consumidor en Internet, es decir, el *nivel de telepresencia y pérdida de la noción del tiempo* cuando se navega. En cambio, el patrón de información en R₁ deja al margen este elemento. En particular, considera las siguientes variables del consumidor: *implicación con la Web, habilidad/control percibido y nivel de atención* en el proceso de navegación y, finalmente, el estado de *flow*. *Grosso modo*, nos dice que niveles intermedios de estos antecedentes explican, con un nivel de confianza de 0,87, estados moderados de comportamientos exploratorios, sin problema inmediato de compra que resolver, en la Web. Brevemente, esta regla que puede inspirar al experto de marketing a considerar nuevos caminos o relaciones no contempladas *a priori*, directamente, entre los antecedentes y consecuente citados. En general, por los antecedentes contenidos en la regla y, en especial, por incluir a *flow* como determinante de comportamientos exploratorios en la Web. Esta relación, que es bastante razonable, siendo contemplada *a priori* en el modelo de base de [4], fue descartada finalmente tras el refinado del modelo. No obstante, reglas como R₁ nos advierten de la conveniencia de su consideración.

R₂: SI implicaciónWeb es {Pequeña o Media} y desafío/estímulo es {Pequeño o Medio} y telepres_distTiempo es Medio y comportamExploratorio es {Medio o Grande} ENTONCES nivelAtención es Medio [Sop: 0,21; Conf: 0,84]

A continuación hemos seleccionado una regla en la que aparece “nivel de atención” como consecuente. En el modelo validado en [4], nivel de atención es una consecuencia de *implicación con la Web* y desafío/estímulo. De nuevo, en la regla R₂ aparecen relaciones no consideradas en [4]. Aparte de *implicación con la Web*

y *desafío/estímulo*, sí contempladas en [4], R_2 sugiere que variables como *telepresencia/pérdida* de la noción del tiempo y *flow* pueden determinar niveles de atención moderados en el proceso de navegación. Esta información, más allá de los niveles tomados por los antecedentes y el consecuente, es interesante por la relación de prelación que establece entre las variables, contraria a la concluida en [4]. Esto es razonable si consideramos el sistema de generación de las reglas de nuestro método. Al contrario que en [4], donde el método estadístico utilizado analiza la procedencia de un modelo establecido por el experto, en nuestro caso el patrón de información que se genera no responde a condicionantes estructurales de ningún tipo. En algunos casos, los patrones obtenidos no tendrán sentido. Sin embargo, no es el caso del patrón en R_2 , que podría justificarse en un proceso dinámico en el que la experiencia del proceso de navegación on-line se retroalimentara cambiando, por tanto, la secuencia de relaciones entre ciertas variables como las analizadas en este caso. En definitiva, R_2 plantea un escenario no contemplado *a priori* por los expertos en [4] que, de haber aplicado un método como el que presentamos en este trabajo, se podría haber estudiado de cara al proceso de diseño y depuración de su modelo.

Más allá de estos ejemplos, pensemos en el soporte que métodos de aprendizaje no supervisado como éste pueden tener para un gestor de marketing, cuando tenga que afrontar problemas de decisión estructurados deficientemente, sin información *a priori*, al proporcionar patrones de información no predefinidos, imprevistos, y en ocasiones inspiradores de soluciones.

5 CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha aplicado Fuzzy-CSar, un sistema clasificador difuso que evoluciona reglas de asociación, para la extracción de modelos casuales en un entorno de marketing. Fuzzy-CSar usa una representación difusa para tratar la incertidumbre de los datos del problema. Se ha comparado el comportamiento del sistema con el presentado por un método multi-objetivo GCCL. Los resultados han mostrado la competitividad de Fuzzy-CSar. Más concretamente, se ha observado empíricamente que Fuzzy-CSar puede conseguir frentes Pareto muy similares a GCCL cuando se fija una variable en el consecuente de las reglas. Además, Fuzzy-CSar tiene la ventaja que puede aprender reglas de asociación sin que previamente se tengan que fijar que variables son de entrada y cuales son de salida. Los experimentos realizados en este entorno han mostrado que Fuzzy-CSar es capaz de extraer reglas de asociación con alto soporte y confianza, dibujando

un frente Pareto muy superior al obtenido con GCCL fijando la variable de salida. Asimismo, Fuzzy-CSar continua generando reglas interesantes para el consecuente *flow*.

Como trabajo futuro, se va a analizar más detalladamente la diversidad de las poblaciones evolucionadas por Fuzzy-CSar. Los resultados presentados indican que el sistema evoluciona una gran diversidad de reglas, de las cuales solo un pequeño porcentaje forman el frente Pareto. Aún así, reglas que estén dominadas por soporte y confianza por otras reglas pueden ser igualmente interesantes si descubren nuevas relaciones entre variables. Por tanto, se investigará más en detalle la semántica de las reglas de la población y se diseñarán metodologías para su comparación.

Referencias

- [1] G. H. V. Bruggen and B. Wierenga. Broadening the perspective on marketing decision models. *International Journal of Research in Marketing*, 17:159–168, 2000.
- [2] T. P. Hong, C. S. Kuo., and S. C. Chi. Trade-off between computation time and number of rules for fuzzy mining from quantitative data. *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 9(5):587–604, 2001.
- [3] F. Martínez-López and J. Casillas. Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic fuzzy systems. *Industrial Marketing Management*, doi:10.1016/j.indmarman.2008.02.003, 2008.
- [4] T. Novak, D. Hoffman, and Y. Yung. Measuring the customer experience in online environments: A structural modelling approach. *Marketing Science*, 19(1):22–42, 2000.
- [5] B. Wierenga and G. H. V. Bruggen. The integration of marketing problem-solving modes and marketing management support systems. *Marketing Management Support Systems*, 61:21–37, 1997.
- [6] B. Wierenga and G. H. V. Bruggen. *Marketing management support systems: principles, tools and implementation*. Kluwer Academic Publishers, 2000.
- [7] M. Zhang. Introduction to data mining and KDD process. Technical report, Victoria University of Wellington. Disponible en: <http://www.mcs.vuw.ac.nz/courses/COMP422/2004T2/Lecturenotes/week01.pdf>, 2004.