

# Un Sistema de Recomendación Basado en Conocimiento con Información Lingüística Multigranular

M. J. Barranco, L.G. Pérez, L. Martínez

Universidad de Jaén

Jaén, España

barranco, lgonzaga, martin@ujaen.es

## Resumen

Los Sistemas de Recomendación son aplicaciones que han surgido en el área del comercio electrónico para ayudar a los usuarios en sus búsquedas en las tiendas electrónicas. Estas tiendas normalmente ofrecen una amplia gama de productos para cubrir las necesidades de una gran variedad de usuarios, sin embargo, para un usuario en concreto puede ser bastante difícil encontrar lo que necesite entre todos los productos ofertados. Los Sistemas de Recomendación ayudan a los usuarios a encontrar lo que necesitan por medio de recomendaciones basadas en la información proporcionada por fuentes tales como otros usuarios, expertos,... La mayoría de los Sistemas de Recomendación actuales obligan a sus usuarios a proporcionar sus preferencias o necesidades en una única escala precisa de información, normalmente numérica y fijada a priori; esto a pesar de que esta información suele ser cualitativa y por tanto su conocimiento incompleto, vago e impreciso. En esta contribución se propone un sistema de recomendación basado en conocimiento que utiliza la aproximación lingüística difusa para modelar la incertidumbre de la información de este tipo de sistemas y que además ofrece un marco de trabajo flexible en el que las necesidades de los usuarios y la descripción de los productos podrán estar valoradas en distintas escalas.

**Palabras clave:** comercio electrónico, servicios electrónicos, sistemas de recomendación, aproximación lingüística difusa, información lingüística multigranular.

## 1. Introducción

Uno de los principales problemas que afrontan los usuarios cuando navegan por Internet es la inmensa cantidad de información que encuentran, siendo la mayoría de ella de poca utilidad. Los usuarios pierden mucho tiempo buscando lo que necesitan, y muchas veces se sienten decepcionados y desisten de sus búsquedas al no encontrar lo que quieren, aún sabiendo que el sistema podría proporcionar una solución a sus necesidades. Diferentes servicios electrónicos han surgido para ayudar a los usuarios en sus procesos de búsqueda. En este trabajo nos centramos en los Sistemas de Recomendación, un tipo de software [Res97] que ha surgido en los últimos años en el área del comercio electrónico [Sch01], y que ayuda a los usuarios a encontrar los productos más adecuados a sus necesidades, preferencias o gustos, ocultando aquellos productos que no le son útiles.

Aunque esencialmente todos los Sistemas de Recomendación tienen el mismo objetivo, guiar al usuario mediante recomendaciones a aquellos productos que más le pueden interesar. Las técnicas utilizadas para llevar a cabo estos objetivos difieren unas de otras significativamente, tanto en la información requerida como en los procesos necesarios para llevar a cabo estas recomendaciones. Basándonos en estos procesos, podemos clasificar los Sistemas de Recomendación en seis tipos:

1. *Sistemas de recomendación demográficos [Kru97]*: clasifican a los usuarios en grupos demográficos basándose en atributos personales y les proporcionan recomendaciones potencialmente interesantes para cualquier persona perteneciente a dicho grupo demográfico.
2. *Sistemas de recomendación basado en contenido [Paz96]*: filtran y recomiendan los productos comparando los términos de consulta de los usuarios y sus actuaciones en el pasado con el índice de términos utilizado en la representación de los productos.
3. *Sistemas de recomendación colaborativo [Gol92]*: utilizan información de preferencia de muchos usuarios para filtrar y recomendar productos a un usuario, ignorando la representación de los objetos. En el caso más simple, estos sistemas predicen las preferencias de un usuario como una suma ponderada de las preferencias de otros usuarios, en los cuales los pesos son proporcionales a las correlaciones sobre el conjunto de objetos comunes evaluados por dos personas.
4. *Sistemas de recomendación basados en conocimiento [Bur00]*: realizan recomendaciones partiendo del conocimiento que da el usuario sobre sus necesidades, y del conocimiento de los productos a recomendar, buscando los que mejor se adapten a las necesidades de los usuarios.
5. *Sistemas de recomendación basados en utilidad [Bur02]*: realizan las recomendaciones basándose en el cálculo de la utilidad de cada objeto con respecto a un perfil de usuario.
6. *Sistemas de recomendación híbridos [Bas98]*: este tipo de sistemas surgió con el objetivo de solventar algunos problemas presentados por los sistemas anteriores ante algunas situaciones. Para ello este tipo de Sistemas de Recomendación combinan diferentes técnicas de funcionamiento de las anteriores.

La información utilizada en los Sistemas de Recomendación suele valorar aspectos, indicadores, parámetros de naturaleza cualitativa porque está relacionada con percepciones, gustos u opiniones de sus usuarios. Por lo que, el conocimiento sobre ellos es vago, impreciso e incompleto. A pesar de esto, la mayoría de ellos obligan a sus usuarios a proporcionar esta información de una forma precisa, usualmente mediante una escala numérica prefijada [Hay01]. Esto implica una pérdida de expresividad y por lo tanto una pérdida de precisión en las recomendaciones.

Esta contribución propone un Sistema de Recomendación basado en conocimiento en el que valoramos las necesidades de los usuarios o las características de los productos mediante la Aproximación Lingüística Difusa [Zad75] que proporciona una metodología que permite modelar de forma directa

información cualitativa como la que utilizan los Sistemas de Recomendación mediante variables lingüísticas. Además este modelo permitirá el uso de diferentes escalas lingüísticas para valorar diferentes necesidades o características dependiendo del conocimiento del usuario definiendo un marco de trabajo *lingüístico multigranular* como los definidos en [Her00]. El uso de información lingüística multigranular implica el uso de operadores adecuados para operar con este tipo de información, ya que, no se puede operar directamente sobre ella. En esta contribución se revisan y presentan los operadores necesarios para realizar los procesos computacionales del Sistema de Recomendación en un marco lingüístico multigranular.

Esta contribución está estructurada de la siguiente forma: en la sección 2 se hace una breve descripción de los Sistemas de Recomendación basados en conocimiento, la sección 3 revisa el enfoque lingüístico difuso, en la sección 4 se introduce un modelo de sistema de recomendación basado en conocimiento con información lingüística multigranular. En la sección 5 se presenta un ejemplo simple que ilustrará el funcionamiento del modelo y finalmente, aportaremos una serie de conclusiones.

## **2. Sistemas de recomendación basados en conocimiento**

Aunque los Sistemas de Recomendación más utilizados y más conocidos son los colaborativos y los basados en contenido, no en todas las situaciones son los más adecuados. Por ejemplo, los Sistemas de Recomendación colaborativos necesitan partir de una base de datos de valoraciones de los usuarios sobre los productos ofertados para poder realizar recomendaciones precisas y acertadas a cualquiera de estos usuarios. Los basados en contenido buscan nuevos productos a recomendar basándose en los valorados por el usuario en el pasado. Por tanto, estos sistemas requieren que el usuario haya valorado un mínimo número de productos para realizar las recomendaciones adecuadas a su proceso de búsqueda de nuevos productos.

En el mundo real nos encontramos situaciones en las que los modelos anteriores no son aplicables por problemas de datos o de tiempo. Para este tipo de situaciones se han presentada varias alternativas para resolver estos problemas, como pueden ser los Sistemas de Recomendación híbridos [Bas98] o los basado en conocimiento [Bur00]. Estos últimos utilizan el conocimiento que proporciona el usuario sobre sus necesidades y el conocimiento que tiene el sistema sobre los productos para desde un enfoque basado en conocimiento, realizar recomendaciones de los productos que mejor cubren las necesidades de los usuarios.

Existen múltiples técnicas para explotar el conocimiento que el usuario aporta sobre sus necesidades y sobre los productos que podemos recomendar. Así por ejemplo, The PersonalLogic utilizaba diálogos para guiar al usuario a través de un árbol de discriminación de las características de los productos hacia los productos más adecuados. Nosotros, en este trabajo nos hemos centrado en los Sistemas de

Recomendación basados en conocimiento que utilizan técnica de razonamiento basado en casos [Ham89] tal y como han utilizado con éxito otros Sistemas de Recomendación como Entree [Bur96, Bur97]. En estos sistemas los usuarios dan un ejemplo del tipo de producto que están buscando, y el sistema buscará y recomendará productos similares al ejemplo dado. Como es de suponer, la mayoría de las veces el usuario lo que busca no es un producto exactamente igual al producto ejemplo si no un producto parecido a este. Por tanto, estos sistemas deben permitir que los usuarios refinen sus búsquedas declarando o modificando algunas de los atributos del ejemplo dado.

### 3. Enfoque lingüístico difuso

Normalmente trabajamos en entornos cuantitativos donde la información se expresa por medio de valores numéricos. Sin embargo, muchos aspectos de diferentes actividades de la vida real no pueden ser valoradas de forma cuantitativa, aunque si de forma cualitativa, i.e., con conocimiento vago e impreciso. En estos casos, un mejor enfoque podría ser la utilización de valoraciones lingüísticas en vez de numéricas. El enfoque lingüístico difuso representa los aspectos cualitativos con valores lingüísticos por medio de variables lingüísticas [Zad75]. Este enfoque es adecuado en situaciones donde la información no puede ser descrita numéricamente debido a su naturaleza cualitativa. Este enfoque ha sido aplicado con éxito en diferentes áreas [Her01, Her03].

Cuando trabajamos con información lingüística hay que seleccionar los descriptores lingüísticos adecuados y su semántica. Para llevar a cabo esta tarea, un aspecto importante a considerar es la *granularidad de la incertidumbre*, es decir, la capacidad de discriminación entre distintos valores de información. Valores típicos de cardinalidad en los modelos lingüísticos son valores impares, tales como 7 ó 9, donde el término medio representa una valoración de "aproximadamente 0,5" y el resto de términos están distribuidos simétricamente a su alrededor. Una posibilidad es generar directamente el conjunto de términos considerando todos los términos distribuidos en una escala sobre la cuál se define un orden total. Un ejemplo de un conjunto de 7 términos es:

$$\{s_0 : N, s_1 : VL, s_2 : L, s_3 : M, s_4 : H, s_5 : VH, s_6 : P\}$$

En estos casos, necesitamos que exista:

- Un operador de negación  $Neg(s_i) = s_j$  tal que  $j = g - i$  ( $g + 1$  es la cardinalidad)
- Un operador de máximo y mínimo en el conjunto de términos lingüísticos:  $s_i \leq s_j \Leftrightarrow i \leq j$

La semántica de los términos se dan por medio de números difusos definidos en  $[0,1]$ . Estos son valoraciones aproximadas, por lo que diferentes autores consideran que una función de pertenencia trapezoidal es lo suficientemente buena para capturar la vaguedad de los términos lingüísticos. Estas

funciones se pueden describir paramétricamente por medio de una 4-tupla  $(a, b, c, d)$ , donde  $b$  y  $d$  indican el intervalo en el cual el valor de pertenencia es 1, con  $a$  y  $c$  indicamos los límites derecho e izquierdo de la definición del dominio de una función de pertenencia trapezoidal [Bon86]. Un caso particular de este tipo de representaciones son las valoraciones lingüísticas cuya función de pertenencia son triangulares, i.e.,  $b = d$ , de esta forma representamos este tipo de función de pertenencia con una 3-tupla  $(a, b, c)$ . En la Figura 1, podemos ver un ejemplo.

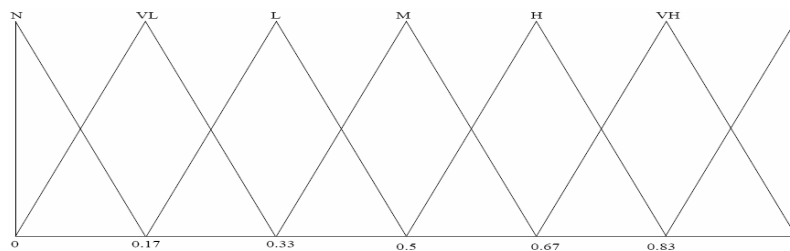


Figura 1: un conjunto de términos lingüísticos

Para operar con información lingüística existen diferentes modelos computacionales en la literatura:

- *Modelo Computacional Semántico* [Deg88]: este modelo opera directamente sobre las funciones de pertenencia de las etiquetas lingüísticas obteniendo como resultados números difusos.
- *Modelo Computacional Simbólico* [Del93]: este modelo opera sobre los índices de las etiquetas lingüísticas obteniendo como resultado valores enteros que corresponderán a índices de etiquetas del conjunto de términos lingüísticos.

#### **4 Modelo de un Sistema de Recomendación Basado en Conocimiento con Información lingüística Multigranular**

A continuación presentamos un modelo para un Sistema de Recomendación basado en conocimiento que solicita al usuario la selección de un caso existente en la base de datos (producto, ítem, ej: *Restaurante Italiano*) para generar los valores de los distintos parámetros o aspectos de los productos o ítems a localizar, estando cada parámetro valorado con una etiqueta lingüística que puede pertenecer a distintos conjuntos de etiquetas. Con este caso ejemplo se genera un perfil inicial que puede ser modificado por el usuario cambiando alguna valoración específica de algún parámetro en particular (ej: Precio: Medio por Muy Alto), pudiendo esta modificación utilizar un conjunto de etiquetas diferente al inicial. Así se genera un perfil final que define las necesidades del usuario y que se utilizará para compararlo con los perfiles de productos de la base de datos para encontrar qué producto cubre mejor las necesidades del usuario.

Este modelo desarrolla su actividad según el siguiente esquema (ver figura 2):

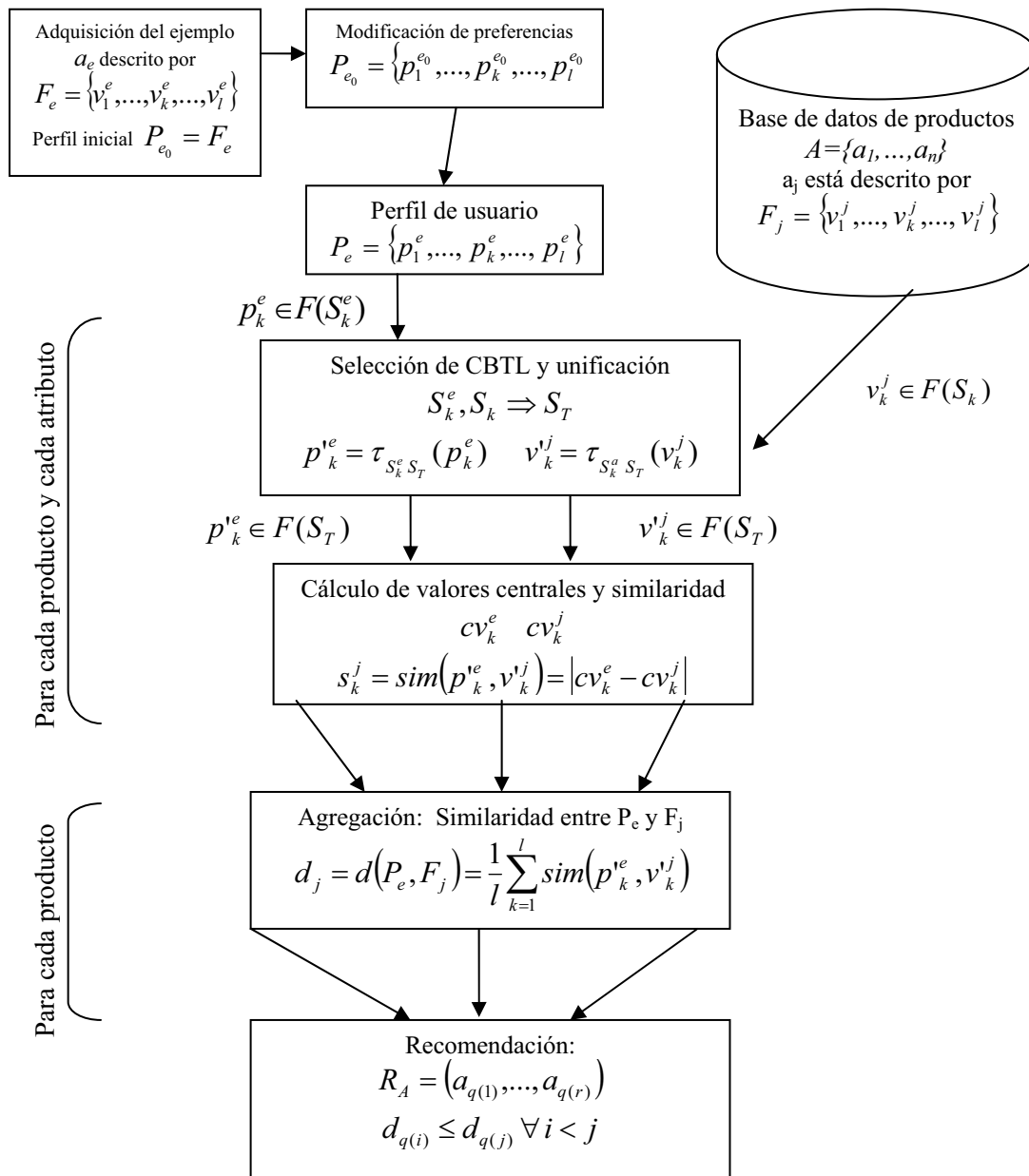


Figura 2: pasos de modelo de sistema de recomendación basado en conocimiento

1. *Obtención del perfil del usuario:* En esta fase se construirá un “perfil de usuario” en el que se almacena la información referente a lo que el usuario está buscando. Esta fase se compone de dos pasos:
  - a. *Adquisición del ejemplo preferido por el usuario:* El usuario selecciona un producto ejemplo de lo que necesita o de lo que le gustaría tener. La descripción existente en la base de datos de este producto definirá el perfil inicial del usuario.

- b. *Modificación de preferencias*: No siempre el usuario busca un producto igual a otro que ya conocía, sino más bien uno parecido que difiera en uno o más atributos. En esta fase el usuario adapta su perfil, utilizando el conjunto de términos lingüísticos más adecuado a su conocimiento y a lo que desea describir, con el fin de refinar el proceso de recomendación.
2. *Cálculo de la similaridad entre el perfil del usuario y los productos*: En esta fase se calcula la satisfacción de las necesidades del usuario por cada producto de la base de datos. Para ello se:
  - a. *Unifica la información lingüística*: Al no poderse operar directamente sobre información lingüística multigranular se unificará en un dominio único que será un conjunto de términos lingüísticos denominado Conjunto Básico de Términos Lingüísticos (CBTL) mediante conjuntos difusos en dicho CBTL.
  - b. *Calculo de la similaridad entre objetos*: para poder recomendar un producto a un usuario, necesitamos conocer la similitud de un producto con las necesidades del usuario. En esta fase calcularemos la similaridad que hay entre cada producto de la base de datos de productos y el perfil del usuario.
3. *Fase de recomendación*: En esta fase se seleccionan los  $r$  productos que mejor satisfacen las necesidades del usuario que serán los recomendados finalmente.

En las siguientes subsecciones presentaremos cada una de estas fases con más detalle.

## 4.1 Obtención del perfil de usuario

En esta fase se obtendrán las necesidades y/o preferencias de un usuario,  $u_e$ , sobre el tipo de producto que busca en nuestro sistema. El sistema de recomendación tendrá una base de datos,  $A = \{a_1, \dots, a_n\}$ , con  $n$  productos cada uno descrito según un conjunto de atributos,  $C = \{c_1, \dots, c_k, \dots, c_l\}$ . Por lo que cada producto  $a_j$ , está descrito por un vector de utilidad  $F_j = \{v_1^j, \dots, v_l^j\}$ , siendo  $v_k^j$  la valoración del atributo  $c_k$  del producto  $a_j$  en el conjunto de etiquetas  $S_k$ . Como vemos cada atributo puede estar valorado en distintos conjuntos de etiquetas dependiendo del conocimiento existente sobre ellos. Una vez que conocemos como se describen los productos en el sistema de recomendación describiremos las fases para obtener el perfil de usuario.

### 4.1.1 Adquisición del ejemplo preferido por el usuario

El punto de partida para definir las necesidades del usuario en este tipo de sistemas se basa en la elección de un ejemplo, es decir, de un producto que el usuario expone como un *caso* de sus necesidades.

Sea  $a_e$  el producto expuesto como ejemplo de las preferencias del usuario  $u_e$ . Este producto estará descrito en la base de datos mediante un vector de utilidad  $F_e = \{v_1^e, \dots, v_l^e\}$ , donde  $v_k^e \in S_k$  es el valor asignado a dicho producto para el atributo  $c_k$  expresado en  $S_k$ .

Este ejemplo seleccionado por el usuario genera un perfil de usuario inicial que notamos como  $P_{e_0} = \{p_1^{e_0}, \dots, p_l^{e_0}\}$ , donde  $p_k^{e_0} = v_k^e$ , ya que se asigna directamente los valores de la base de datos a los del perfil. En este perfil inicial los conjuntos de términos lingüísticos son los mismos que los utilizados en la base de datos.

#### 4.1.2 Modificación ocasional de preferencias

Una vez definido el perfil de usuario inicial se ofrece al usuario la posibilidad de alterar algún valor de su perfil dando una valoración para un atributo determinado expresado en un conjunto de términos lingüísticos a su elección o en el ya existente. En tal caso, para un atributo  $c_k$ , el usuario podrá asignarle una nueva valoración,  $p_k^{e_1}$ , expresada en otro conjunto de términos lingüísticos,  $S_k'$ , más acorde con su conocimiento.

Después de esta modificación que es opcional tenemos un perfil definitivo de usuario  $P_e = \{p_1^e, \dots, p_l^e\}$  donde  $p_k^e \in S_k^e$  obtenidos del siguiente modo:

- $p_k^e = p_k^{e_0}$ ,  $p_k^e \in S_k^e = S_k$  si el atributo  $c_k$  no ha sido modificado
- $p_k^e = p_k^{e_1}$ ,  $p_k^e \in S_k^e = S_k'$  en caso contrario.

En esta fase de nuestro modelo se ofrece a los usuarios la posibilidad de describir sus necesidades utilizando sus propios conjuntos de términos lingüísticos acordes con su conocimiento, sin necesidad de tener que adaptarse a una escala única fijada a priori.

#### 4.2 Cálculo de la Similitud entre el Perfil del Usuario y los Productos

Una vez hemos construido el perfil de usuario,  $P_e = \{p_1^e, \dots, p_l^e\}$ , el sistema ha de encontrar qué productos de  $A = \{a_1, \dots, a_n\}$  son los más adecuados de acuerdo a dicho perfil.

El objetivo de esta fase es medir la similitud de los productos a las necesidades de usuario. Para ello se realizan los siguientes pasos:



1. Unificar la información lingüística: Al trabajar con información lingüística multigranular no podemos operar directamente sobre ella por lo que se unifica en un único dominio de expresión mediante conjuntos difusos.
2. Calcular la distancia entre cada producto y el perfil de usuario: Una vez unificada la información se mide la similaridad entre el perfil de usuario y los productos.

#### 4.2.1 Unificar la información lingüística

Para manejar la información multigranular debemos unificarla en un único dominio de expresión [Her00]. En este caso, hemos escogido como dominio de unificación un conjunto básico de términos lingüísticos o CBTL y representado por  $S_T$ . La información será unificada mediante conjuntos difusos en el CBTL,  $F(S_T)$ .

Antes de definir ninguna función de transformación debemos establecer cómo escoger el  $S_T$ . El CBTL debe ser un conjunto que permita mantener, tanto el grado de incertidumbre como la capacidad de discriminación de cada valor en el marco de definición del problema. Con este objetivo, se busca un CBTL de máxima granularidad. Si hay varias opciones seguiremos las siguientes reglas [Her00]:

1. Cuando todos los conjuntos de términos lingüísticos tienen la misma semántica, entonces escogemos cualquiera de ellos como  $S_T$ .
2. Si hay algunos conjuntos de términos lingüísticos con diferente semántica entonces  $S_T$  será un conjunto de términos lingüísticos con un número de términos más grande que el número de términos que una persona es capaz de discriminar (normalmente 11 o 13 [Mil56]).

Una vez hayamos seleccionado el CBTL pasaremos a expresar los términos lingüísticos de los distintos conjuntos de etiquetas en conjuntos difusos definidos en el CBTL,  $F(S_T)$ , mediante la siguiente función de transformación.

**Definición 1.-** [Her00] Sean  $A = \{l_0, \dots, l_p\}$  y  $S_T = \{s_0, \dots, s_g\}$  dos conjuntos de términos lingüísticos tales que  $g \geq p$ . Entonces, una función de transformación multigranular,  $\tau_{AS_T}$  esta definida como

$$\tau_{AS_T} : A \rightarrow F(S_T)$$

$$\tau_{AS_T}(l_i) = \left\{ (s_k, \alpha_k^i) / k \in \{0, \dots, g\} \right\}, \forall l_i \in A$$

$$\alpha_k^i = \max_y \min \{ \mu_{l_i}(y), \mu_{s_k}(y) \}$$

donde  $F(S_T)$  es el conjunto de todos los conjuntos difusos definidos en  $S_T$ , y  $\mu_{l_i}(y)$  y  $\mu_{s_k}(y)$  son funciones de pertenencia de los conjuntos difusos asociados a los términos  $l_i$  y  $s_k$  respectivamente.

**Nota:** En caso de que el conjunto de términos lingüísticos,  $A$ , sea el escogido como CBTL, el conjunto difuso que representa un término lingüístico tendrá todos sus elementos a 0 excepto el valor correspondiente al ordinal de la etiqueta lingüística, que será 1.

**Ejemplo:** Sea  $A = \{l_0, \dots, l_8\}$ , y sea  $S_T = A$ , si queremos expresar la etiqueta lingüística  $l_2$  en  $F(S_T)$  obtendremos que:  $\tau_{AS_T}(l_2) = (0, 0, 1, 0, 0, 0, 0, 0, 0)$ .

Dado que cada atributo,  $c_k$ , de los productos que podemos recomendar esta valorado en el conjunto de etiquetas  $S_k$ , y dado que en el perfil de usuario para valorar dicho atributo podemos emplear otro conjunto de etiquetas  $S'_k$  distinto del inicial y más acorde con el conocimiento del usuario, para poder explotar toda esta información y realizar las recomendaciones necesitaremos unificar todo esta información en un mismo dominio, el CBTL. Para llevar a cabo estas transformaciones emplearemos tantas funciones  $\tau_{S'_k S_T}$  y  $\tau_{S_k S_T}$  como sean necesarias, obteniendo para cada componente del perfil de usuario  $p_k^e$  un conjunto difuso definido sobre el CBTL,  $p_k^e$  que representaremos mediante los valores de sus grados de pertenencia  $(\alpha_{k1}^e, \dots, \alpha_{kg}^e)$  y para cada atributo  $c_k$  de cada producto  $a_j$  tendremos que su valor  $v_k^j$  se transforma en un conjunto difuso  $v_k^j$  que representaremos de igual manera que el anterior  $(\alpha_{k1}^j, \dots, \alpha_{kg}^j)$ . Una vez expresados en el mismo dominio de expresión, podemos proceder a calcular la distancia entre ambos.

#### 4.2.2 Calculo de la Similitud entre el Perfil de Usuario y los Productos

Una vez unificada la información el sistema debe descubrir qué productos satisfacen mejor las necesidades del usuario. Para ello se necesita calcular la similitud entre vectores de conjuntos difusos que representan, por un lado, el perfil del usuario, y por otro lado cada uno de los productos del sistema. Una vez calculado el grado de similitud entre el perfil del usuario y los productos, el sistema puede conocer cómo satisface cada producto las necesidades del usuario.

Tanto el perfil de usuario como los productos están descritos por vectores de conjuntos difusos donde cada conjunto describe un atributo. Para calcular la similitud entre el perfil y un producto necesitamos conocer la distancia entre las valoraciones de cada atributo. Inicialmente pensamos en las funciones de distancias clásicas, pero las deseamos, después de comprobar que los resultados no eran los esperados como se muestra en el siguiente ejemplo.

**Ejemplo.** Supongamos que después de la unificación de la información hemos obtenido para el atributo  $c_k$  los siguientes conjuntos difusos  $p_k^e = (1, 0, 0, 0, 0, 0, 0)$ ,  $v_k^i = (0, 0, 0, 1, 0, 0, 0)$  y  $v_k^j = (0, 0, 0, 0, 0, 1, 0)$ ,

donde  $p_k^{ie}$  (inicialmente *Nulo*) corresponde al valor que en el perfil de usuario se otorga al atributo  $c_k$ , y  $v_k^{ii}$  (inicialmente *Medio*) y  $v_k^{jj}$  (inicialmente *Alto*) son las valoraciones para los productos  $a_i$  y  $a_j$  respectivamente para el mismo atributo. Utilizando la distancia euclídea para comparar estos valores obtenemos los siguientes resultados:

$$ED(p_k^{ie}, v_k^{ii}) = \sqrt{\sum_{h=0}^g (p_{k_h}^{ie} - v_{k_h}^{ii})^2} = \sqrt{1+0+0+1+0+0+0} = \sqrt{2} = 0.41$$

$$ED(p_k^{ie}, v_k^{jj}) = \sqrt{\sum_{h=0}^g (p_{k_h}^{ie} - v_{k_h}^{jj})^2} = \sqrt{1+0+0+0+0+1+0} = \sqrt{2} = 0.41$$

Vemos que ambas valoraciones  $v_k^{ii}$  y  $v_k^{jj}$  están a la misma distancia de la valoración  $p_k^{ie}$ , aunque, si las interpretamos lingüísticamente (*Nulo*, *Medio*, *Alto*) la segunda está más lejos que la primera. El problema, en este caso, es la forma en que la distancia interpreta los conjuntos difusos, ya que los considera como vectores de grados de pertenencia sin tener en cuenta sus posiciones en él. Sin embargo para obtener valores correctos en el cálculo de la similaridad, la posición ha de tomarse en consideración. Para hacer esto, definimos una función que basa el cálculo de distancias en el valor central,  $cv$ , de conjuntos difusos. Dado un conjunto difuso  $b' = (\alpha_1, \dots, \alpha_g)$  definido sobre  $S = \{s_h\}$  para  $h=0..g$ , obtenemos su valor central  $cv$  del siguiente modo:

$$cv = \frac{\sum_{h=0}^g index(s_h) \alpha_h}{\sum_{h=0}^g \alpha_h}, \text{ donde } index(s_h) = h$$

que representará la posición media o centro de gravedad de la información contenida en el conjunto difuso  $b'$ . El rango de este valor central es el intervalo cerrado  $[0, g]$ .

**Ejemplo:** Si aplicamos esta fórmula a las valoraciones del ejemplo anterior obtendremos los siguientes valores centrales:

$$cv_k^{P_e} = 0, \quad cv_k^{a_i} = 3, \quad cv_k^{a_j} = 5$$

Como puede verse, de esta forma la valoración de  $v_k^{jj}$  está más lejos que  $v_k^{ii}$  de la valoración empleada en el perfil de usuario,  $p_k^{ie}$ . Si consideremos ahora los conjuntos difusos  $b'_1 = (0.3, 0.8, 0.6, 0, 0, 0)$ ,  $b'_2 = (0, 0.3, 0.8, 0.6, 0, 0)$ , y  $b'_3 = (0, 0, 0, 0.3, 0.8, 0.6)$ . Aplicando la fórmula anterior obtenemos como valores centrales:

$$cv_1 = 1.18, \quad cv_2 = 2.18, \quad cv_3 = 4.18$$

Como esperábamos, cuando la información (valores de pertenencia) se mueven desde la parte izquierda del conjunto difuso a la parte derecha, el valor central se incrementa.

Basándonos en los cálculos de los valores centrales podemos definir la siguiente medida de similaridad que emplearemos para calcular la similaridad entre cada atributo,  $c_k$ , del perfil de usuario  $P_e$  y de un producto  $a_j$

**Definición 2.-** Sean  $b_1'$  y  $b_2'$  dos conjuntos difusos definidos sobre el CBTL,  $S_T = \{s_0, \dots, s_g\}$ , y sean  $cv_1$  y  $cv_2$  los valores centrales de  $b_1'$  y  $b_2'$  respectivamente, entonces la similaridad entre ellos es calculada como:

$$sim(b_1', b_2') = |cv_1 - cv_2|$$

Si utilizamos esta función, podemos calcular la similaridad entre el perfil de usuario  $P_e$  y un producto  $a_j$  siguiendo los siguientes pasos:

1. Para cada atributo  $c_k$  calculamos la similaridad entre  $p_k^e$  y  $v_k^j$ :
  - a. Obtenemos el valor central  $cv_k^{P_e}$  del conjunto difuso  $p_k^e$ , correspondiente al objeto  $P_e$  y el valor central  $cv_k^{a_j}$  del conjunto difuso  $v_k^j$  para el objeto  $a_j$ .
  - b. Calculamos su similaridad según la función de la Def. 2:

$$sim(p_k^e, v_k^j) = |cv_k^{P_e} - cv_k^{a_j}|$$

Como resultado de este primer paso, una vez realizado para los  $l$  atributos, obtenemos un vector que indica la similaridad de cada atributo entre  $P_e$  y  $a_j$ :

$$V\_SIM(P_e, a_j) = \{sim(p_1^e, v_1^j), \dots, sim(p_k^e, v_k^j), \dots, sim(p_l^e, v_l^j)\}$$

2. Para obtener la similaridad total,  $d_j$ , entre  $P_e$  y  $a_j$  agregamos las similaridades de  $V\_SIM(P_e, a_j)$  mediante un operador de agregación. Si todos los atributos tienen la misma importancia puede utilizarse la media aritmética tal y como se muestra a continuación:

$$d_j = d(P_e, a_j) = \frac{1}{l} \sum_{k=1}^l sim(p_k^e, v_k^j), \quad sim(p_k^e, v_k^j) \in V\_SIM(P_e, a_j)$$

### 4.3 Recomendación

Una vez calculada la similaridad entre el perfil de usuario y todos los objetos de la base de datos pueden ordenarse los productos en base a las similaridades obtenidas y que están representadas en el siguiente vector de similaridad  $D=(d_1, \dots, d_n)$ . Los mejores serán aquellos que mejor satisfagan las necesidades del perfil del usuario (con la similaridad más pequeña). Así pues, dado el conjunto de productos  $A=\{a_1, \dots, a_n\}$  y dado un número  $r$  de productos a recomendar, la recomendación realizada al usuario vendrá dada por el siguiente vector de recomendación,  $R_A$ , siendo el primer elemento el más recomendado:

$$R_A=(a_{q(1)}, \dots, a_{q(r)})$$

donde la función  $q$  se define del siguiente modo:

- $q : \{1, 2, \dots, r\} \rightarrow \{1, 2, \dots, n\}$
- $q(i) \neq q(j) \quad \forall i \neq j$
- $q(i) \neq e \quad \forall i = 1..r$ , siendo  $a_e$  el ejemplo dado por el usuario
- $d_{q(i)} \leq d_{q(j)} \quad \forall i < j$
- Si  $d_k < d_{q(r)}$  y  $k \neq e$  entonces  $\exists m < r \mid q(m) = k, \forall k = 1..n$

### 5 Ejemplo

Para simplificar nuestro ejemplo hemos partido de que nuestro sistema de recomendación tiene una base de datos de productos  $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$  descritos por un conjunto de atributos  $C = \{c_1, c_2, c_3, c_4\}$ . En sistemas reales estas bases de datos podrían tener almacenados cientos o miles de productos y estarán descritos por más atributos. Para describir los atributos  $c_1$  y  $c_2$  hemos empleado el conjunto de etiquetas  $S_{1,2}$  (ver figura 3) y para describir los atributos  $c_3$  y  $c_4$  hemos empleado el conjunto de etiquetas  $S_{3,4}$  (ver figura 4). Estos conjuntos lingüísticos los definiremos mediante las siguientes funciones de pertenencia:

$$\begin{array}{lll}
 s_0^{1,2} = \text{Infimo} = (0, 0, 0, 0.125) & s_1^{1,2} = \text{Muy bajo} = (0, 0.125, 0.25) & s_2^{1,2} = \text{Bajo} = (0.125, 0.25, 0.375) \\
 s_3^{1,2} = \text{Un poco bajo} = (0.25, 0.375, 0.5) & s_4^{1,2} = \text{Medio} = (0.375, 0.5, 0.625) & s_5^{1,2} = \text{Un poco alto} = (0.5, 0.75, 0.625) \\
 s_6^{1,2} = \text{Alto} = (0.625, 0.75, 0.875) & s_7^{1,2} = \text{Muy alto} = (0.75, 0.875, 1) & s_8^{1,2} = \text{Extremo} = (0.875, 1, 1)
 \end{array}$$

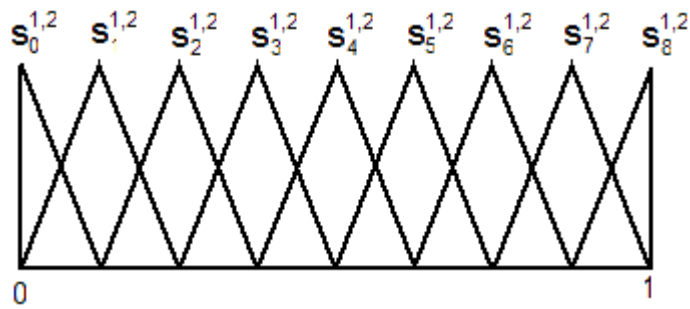


Figura 3: Conjunto de etiquetas  $S_{1,2}$  empleadas en la base de datos de productos.

$s_0^{3,4} = \text{Despreciable} = (0,0,0.16)$    
  $s_1^{3,4} = \text{Muy Inferior} = (0,0.16,0.33)$    
  $s_2^{3,4} = \text{Inferior} = (0.16,0.33,0.5)$   
 $s_3^{3,4} = \text{Normal} = (0.33,0.5,0.66)$    
  $s_4^{3,4} = \text{Elevado} = (0.5,0.66,0.83)$    
  $s_5^{3,4} = \text{Muy Elevado} = (0.66,0.83,1)$   
 $s_6^{3,4} = \text{Considerable} = (0.83,1,1)$

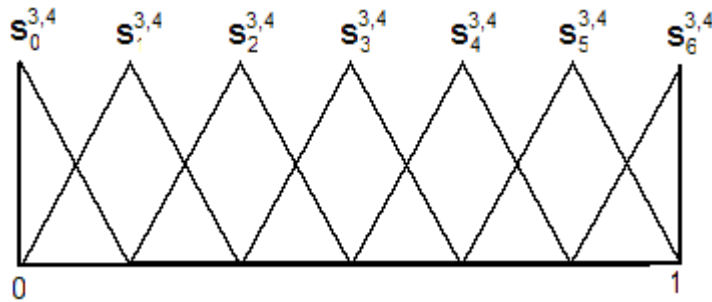


Figura 4: Conjunto de etiquetas  $S_{3,4}$  empleadas en la base de datos

Las descripciones de los productos de la base de datos las podemos ver en la tabla 1

	$c_1$	$c_2$	$c_3$	$c_4$
$a_1$	$s_0^{1,2}$	$s_3^{1,2}$	$s_3^{3,4}$	$s_2^{3,4}$
$a_2$	$s_5^{1,2}$	$s_2^{1,2}$	$s_1^{3,4}$	$s_4^{3,4}$
$a_3$	$s_7^{1,2}$	$s_4^{1,2}$	$s_0^{3,4}$	$s_5^{3,4}$
$a_4$	$s_5^{1,2}$	$s_6^{1,2}$	$s_2^{3,4}$	$s_6^{3,4}$
$a_5$	$s_8^{1,2}$	$s_0^{1,2}$	$s_4^{3,4}$	$s_6^{3,4}$
$a_6$	$s_1^{1,2}$	$s_8^{1,2}$	$s_3^{3,4}$	$s_1^{3,4}$

Tabla 1: Base de datos de productos

Un usuario  $u_e$  desea recibir una recomendación de nuestro sistema. Lo primero que tenemos que hacer es **obtener el perfil del usuario**. Para ello debemos seguir los siguientes pasos:

1. *Adquisición del ejemplo preferido por el usuario:* el producto  $a_1$  es el ejemplo que el usuario nos ha dado porque es un producto conocido por él, que le gustó y que es el más cercano a las necesidades actuales. Con esta información construiremos el siguiente perfil inicial del usuario:

$$P_e = \{s_0^{1,2}, s_3^{1,2}, s_3^{3,4}, s_2^{3,4}\}$$

2. *Modificación ocasional de preferencias:* aunque este ejemplo representa bastante bien las necesidades del usuario, este quiere refinar su búsqueda cambiando alguno de estos valores. El valor del atributo  $c_1$  quiere aumentarlo pero el conjunto de etiquetas utilizado por el sistema para describir este atributo es demasiado preciso para el nivel de conocimiento que tiene. El usuario decide utilizar el conjunto de etiquetas  $S'_1$  que definimos a continuación:

$$s_0^{1l} = \text{Muy alto} = (1,1,0.75) \quad s_2^{1l} = \text{Medio} = (0.75,0.5,0.25) \quad s_4^{1l} = (0.25,0,0)$$

$$s_1^{1l} = \text{Alto} = (1,0.75,0.5) \quad s_3^{1l} = \text{Bajo} = (0.5,0.25,0)$$

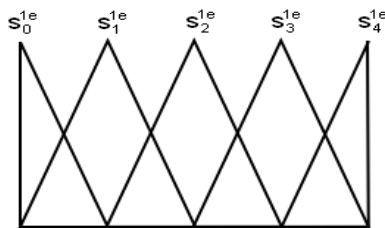


Figura 5: Conjunto de etiquetas utilizado por el usuario

El usuario le asigna el valor  $s_2^{1l}$  al atributo  $c_1$ . Por lo tanto, después de este paso su perfil de usuario será el siguiente:

$$P_e = \{s_2^{1l}, s_3^{1,2}, s_3^{3,4}, s_2^{3,4}\}$$

Como podemos comprobar ahora el perfil de usuario está definido en un contexto multigranular pues utilizamos etiquetas pertenecientes a distintos conjuntos de términos lingüísticos y cada uno de ellos con una granularidad distinta.

El siguiente paso en nuestro ejemplo es el **cálculo de la distancia entre el perfil de usuario y los productos de la base de datos de productos**. Para ello debemos realizar los siguientes pasos:

1. *Unificar la información lingüística:* En este paso unificaremos la información del perfil del usuario y de la descripción de los productos en un único dominio. Si revisamos las condiciones enunciadas en 4.2.1 para la elección del CBTL nos daremos cuenta que el conjunto  $S_{1,2}$  es un conjunto candidato para ser el CBTL  $S_T$ . Si aplicamos la función de transformación al perfil de usuario y a la base de datos de productos obtendremos el siguiente perfil de usuario:

$$P_e = \{(0,0,0.33,0.66,1,0.66,0.33,0,0), (0,0,0,1,0,0,0,0), (0,0,0.14,0.57,1,0.57,0.14,0,0), (0,0.28,0.71,0.85,0.42,0,0,0)\}$$

Y los siguientes valores para la descripción de los productos de la base de datos de productos:

	$c_1$	$c_2$	$c_3$	$c_4$
$a_1$	(1,0,0,0,0,0,0,0)	(0,0,0,1,0,0,0,0)	(0,0,.14,.57,1,.57,.14,0,0)	(0,.28,.71,.85,.42,0,0,0)
$a_2$	(0,0,0,0,0,1,0,0)	(0,0,1,0,0,0,0,0)	(.42,.85,.71,.28,0,0,0,0)	(0,0,0,0,.42,.85,.71,.28,0)
$a_3$	(0,0,0,0,0,0,0,1)	(0,0,0,0,1,0,0,0)	(1,.57,.14,0,0,0,0,0)	(0,0,0,0,.28,.71,.85,.42)
$a_4$	(0,0,0,0,0,1,0,0)	(0,0,0,0,0,0,1,0)	(0,.28,.71,.85,.42,0,0,0)	(0,0,0,0,0,.14,.57,1)
$a_5$	(0,0,0,0,0,0,0,1)	(1,0,0,0,0,0,0,0)	(0,0,0,0,.42,.85,.71,.28,0)	(0,0,0,0,0,.14,.57,1)
$a_6$	(0,1,0,0,0,0,0,0)	(0,0,0,0,0,0,0,1)	(0,0,.14,.57,1,.57,.14,0,0)	(.42,.85,.71,.28,0,0,0,0)

Tabla 2: Descripciones de los productos expresados en el CBTL

Para mostrar como se han obtenido estos valores a continuación mostraremos los pasos seguidos para expresar el atributo  $c_1$  del perfil de usuario al CBTL siendo  $S_T = S_{1,2}$  :

$$\tau_{s_1, S_T}(s_2^{n1}) = \{(s_0^T, 0), (s_1^T, 0), (s_2^T, 0.33), (s_3^T, 0.66), (s_4^T, 1), (s_5^T, 0.66), (s_6^T, 0.33), (s_7^T, 0), (s_8^T, 0)\} = (0,0,0.33,0.66,1,0.33,0,0)$$

2. *Calculo de la distancia entre dos objetos:* En esta fase calcularemos la distancia entre el perfil de usuario y el resto de objetos. Para ello seguiremos los siguientes pasos:
  - a. Cálculo de los valores centrales tanto del perfil de usuario como de los productos de la base de datos de productos

$$P_e^{VC} = \{4,3,4,2.62\}$$

	$c_1$	$c_2$	$c_3$	$c_4$
$a_1$	0	3	4	2.62
$a_2$	5	2	1.37	5.37
$a_3$	7	4	1.19	7.05
$a_4$	5	6	2.62	7.5
$a_5$	8	0	5.37	7.5
$a_6$	1	8	4	1.37

Tabla 3: Valores centrales tanto del perfil de usuario como de los productos



A continuación, a modo de ejemplo, mostraremos los pasos seguidos para obtener el valor central del atributo  $c_1$  del perfil de usuario:

$$cv(0,0,0.33,0.66,1,0.66,0.33,0,0) = \frac{0 \cdot 0 + 1 \cdot 0 + 2 \cdot 0.33 + 3 \cdot 0.66 + 4 \cdot 1 + 5 \cdot 0.66 + 6 \cdot 0.33 + 7 \cdot 0 + 8 \cdot 0}{0 + 0 + 0.33 + 0.66 + 1 + 0.66 + 0.33 + 0 + 0} = 4$$

- b. Cálculo de la distancia entre los atributos del perfil de usuario y la base de datos de productos:

	$c_1$	$c_2$	$c_3$	$c_4$
$a_1$	4	0	0	0
$a_2$	1	1	2.63	2.75
$a_3$	3	1	2.81	4.43
$a_4$	1	3	1.38	4.88
$a_5$	4	3	1.37	4.88
$a_6$	3	5	0	1.25

Tabla 4: Distancia entre los atributos de la base de datos de productos y el perfil de usuario

Para mostrar como se han obtenido estos valores calcularemos la similitud del atributo  $c_1$  entre el perfil de usuario y el producto  $a_1$ :

$$sim(p_1^e, v_1^u) = |cv_1^{p_e} - cv_1^{a_1}| = |4 - 0| = 4$$

- c. El último paso de esta fase es el cálculo de la distancia entre los productos de la base de datos y el perfil de usuario obteniendo los siguientes resultados:

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
1	1.84	2.81	2.57	3.31	2.31

Tabla 5: Distancia entre los productos y el perfil de usuario

A continuación mostraremos como se calcularía la distancia entre el perfil de usuario y el producto  $a_1$ :

$$d_1 = d(P_e, a_1) = \frac{4 + 0 + 0 + 0}{4} = 1$$

La última fase que nos queda por completar es la fase de **recomendación**. En esta fase debemos recomendar aquellos productos que más se acerquen a las necesidades del usuario. Si nos damos cuenta el producto que más se acerca al perfil de usuario es  $a_1$  lo cual es bastante lógico teniendo en cuenta que este fue el ejemplo que nos dio el usuario. Por esta razón,  $a_1$  no puede ser recomendado, ya que el

usuario está buscando productos similares a  $a_1$  pero que no sean  $a_1$ . Por lo tanto el vector ordenado que obtenemos es  $\{a_2, a_6, a_4, a_3, a_5\}$ . Si partimos de que nuestro sistema de recomendación solo recomienda los dos productos más cercanos, recomendaríamos:

$$\{a_2, a_6\}$$

## 6 Conclusiones

En esta contribución hemos presentado un modelo de sistema de recomendación basado en conocimiento. Este modelo de sistema de recomendación explota la información disponible sobre el usuario empleando técnicas de razonamiento basado en casos. Este tipo de Sistemas de Recomendación se caracterizan no solo por la forma en la que explotan el conocimiento disponible, sino también porque requieren menor información que otro tipo de Sistemas de Recomendación como pueden ser los colaborativos o los basados en contenido. Además, este sistema de recomendación ofrece un contexto de trabajo más adecuado que los clásicos pues maneja información lingüística en vez de numérica para recoger la información sobre los gustos, necesidades u opiniones de los usuarios, permitiéndoles a éstos no solo emplear información lingüística, sino también que se expresen con los conjuntos de términos lingüísticos más adecuados al conocimiento que ellos tienen sobre el atributo que están valorando. Así mismo, a diferencia de otras contribuciones, nosotros hemos propuesto el empleo de una medida de distancia para calcular no solo la coincidencia o semejanza de dos valores, sino también para conocer cuanto difieren dichos valores unos de otros, lo cual nos permite mejorar la precisión de nuestras recomendaciones.

## Agradecimientos

Esta investigación ha sido subvencionada por el proyecto de investigación *Soft Modelling in Decision Making. On-line Applications (II): Recommendation Systems* (MTM-2005-08982-C04-03) y por la Consejería de Turismo, Comercio y Deporte de la Junta de Andalucía.

## Bibliografía

- [Bas98] C. Basu, H. Hirsh, and W. Cohen. *Recommendation as classification: Using social and contentbased information in recommendation*. Proceedings of the Fifteenth National Conference on Artificial Intelligence, 714-720. 1998.
- [Bon86] P. Bonissone and K. Decker. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity*. In: *Uncertainty in Artificial Intelligence* (Kanal, L. N.; Lemmer, J. F., eds.), 217-247. Amsterdam: North-Holland. 1986.

- [Bur96] R. Burke, K. Hammon, and E. Cooper. *Knowledge-based navigation of complex information spaces*. In proceedings of the 13<sup>th</sup> National Conference on Artificial Intelligence, pp. 462-468. Menlo Park, CA:AAAI Press. 1996.
- [Bur97] R. Burke, K. Hammond, and B. Young. *The FIndMe Approach to Assisted Browsing*. IEEE Expert, 12(4), pp. 32-40. 1997.
- [Bur00] R. Burke. *Knowledge-based Recommender Systems*. In A. Kent (ed.), Encyclopedia of Library and Information Systems. Vol. 69, Supplement 32. New York: Marcel Dekker. 2000.
- [Bur02] R. Burke. *Hybrid Recommender Systems: Survey and Experiments*. User Modeling and User-Adapted Interaction, 12(3), pp. 331-370. 2002.
- [Deg88] R. Degani, G. Bortolan. *The problem of linguistic approximation in clinical decision making*. International Journal of Approximate Reasoning 2, pp. 143–162. 1988.
- [Del93] M. Delgado, J.L. Verdegay, M.A. Vila, *On aggregation operations of linguistic labels*. International Journal of Intelligent Systems 8, pp. 351–370. 1993.
- [Gol92] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry. *Using collaborative filtering to weave an information tapestry*. Communications of the ACM CACM 35(12), 61- 70. 1992.
- [Ham89] K. Hammond. *Case-base planning: Viewing Planning as a Memory Task*. Boston, MA: Academic Press. 1989.
- [Hay01] C. Hayes and P. Cunningham. *Smart radio - community based music radio*. Knowledge-Based Systems, 14(3-4), pp. 197-201. 2001.
- [Her00] F. Herrera, E. Herrera-Viedma, L. Martínez. *A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making*. Fuzzy Sets and Systems. 114, pp. 43-58. 2000.
- [Her01] E. Herrera-Viedma. *Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach*. J. of the American Society for Information Science and Technology. 52(6), 460-475. 2001
- [Her03] E. Herrera-Viedma, E. Peis. Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. Information Processing & Management, 39(2) , pp. 195-213. 2003.
- [Kru97] B. Krulwich. *Lifestyle Finder: Intelligent user profiling using large-scale demographic data*. AI Magazine, 18(2), pp. 37-45. 1997.
- [Mil56] G. Miller. *The magical number seven or minus two: Some limits on our capacity of processing information*. Psychological Review, (63), pp. 81-97. 1956.
- [Paz96] M. Pazzani, J. Muramatsu, and D. Billsus. *Syskill and Webert: Identifying interesting web sites*. In Proceedings of the Thirteenth National Conference on Artificial Intelligence AAAI, 96, pp. 54-61. 1996.

- [Res97] P. Resnick and H.R. Varian. *Recommender systems*. Association for Computing Machinery. Communications of the ACM., 40(3), pp. 56. Mar 1997.
- [Sch01] J.B. Schafer, J.A. Konstan, and J. Riedl. *E-commerce recommendation applications*. Data Mining and Knowledge Discovery, (5), pp. 115-153. 2001.
- [Zad75] L.A. Zadeh. *The cocept of a linguistic variable and its application to approximate reasoning*. Information Science, (8 and 9):(Part I and II) 8, pp. 199-249 and pp. 301-357, (Part III) 9, pp. 43-80, 1975.