

# Estableciendo una medida de *Trust* para seleccionar fuentes de información relevantes para recomendar

**Silvana Vanesa  
Aciar**

Dept. d'Electrònica,  
Informàtica i  
Automàtica  
Universitat de Girona  
Campus Montilivi  
17071 Girona  
[saciar@eia.udg.es](mailto:saciar@eia.udg.es)

**Javier Guzmán  
Obando**

Dept. d'Electrònica,  
Informàtica i  
Automàtica  
Universitat de Girona  
Campus Montilivi  
17071 Girona  
[jguzmano@eia.udg.es](mailto:jguzmano@eia.udg.es)

**Josep Lluís de la  
Rosa**

Dept. d'Electrònica,  
Informàtica i  
Automàtica  
Universitat de Girona  
Campus Montilivi  
17071 Girona  
[pepluis@eia.udg.es](mailto:pepluis@eia.udg.es)

**Josefina López  
Herrera**

Dept. d'Electrònica,  
Informàtica i  
Automàtica  
Universitat de Girona  
Campus Montilivi  
17071 Girona  
[josefina.lopez@udg.es](mailto:josefina.lopez@udg.es)

## Resumen:

Gestionar eficiente y eficazmente la información de los clientes o de los consumidores en mercados masivos es una condición necesaria para la implementación de procesos CRM o de un marketing de relaciones. Las tecnologías de la información aplicadas sobre el conjunto de datos han de generar un análisis del comportamiento del cliente o usuario del sistema de intercambio, sintetizando una información que facilitará y mejorará la personalización en relación a sus procesos de compra. Formando parte de las tecnologías de la información destacan los Sistemas Recomendadores. Los sistemas recomendadores adquieren las preferencias de los consumidores y le presenta los productos que ellos desean. En el presente trabajo se presenta un conjunto de características intrínsecas de las fuentes que junto con una medida de *trust* de fuentes similares permiten conocer a priori la eficacia de un sistema recomendador.

## 1. Introducción

Tener y saber obtener información sobre los clientes o sobre los consumidores en mercados masivos es fundamental para el desarrollo de relaciones con los clientes o consumidores. En esta cuestión, las tecnologías de la información, cuanto más modernas y avanzadas mejor, surgen como el socio o compañero estratégico y/o operativo imprescindible para poder implementar la estrategia de relaciones y convertir al marketing masivo de las organizaciones y su conocimiento de los consumidores en lo más parecido posible a los antiguos tenderos. Se trata, por tanto, de un problema de cantidad, de número. Si el número de clientes no es excesivamente elevado es manejable la información necesaria para desarrollar unas buenas relaciones y si el número empieza a crecer es cuando hay que emplear las tecnologías de la información ofreciendo una información de calidad sobre los consumidores. Consumidores que se cuentan por miles en los mercados de productos de gran consumo.

Es precisamente al nivel de contacto con el cliente donde los sistemas recomendadores pueden emplearse para incrementar la eficiencia en el manejo de la información y la eficacia en los contactos. El recomendador es una aplicación que proponen productos y

proveen información para facilitar la toma de decisiones al consumidor [8]. Un sistema recomendador es una herramienta que toma como entradas las preferencias y comportamiento de los usuarios e información de los productos o servicios y genera recomendaciones de aquellos que le gustaría más a los usuarios [8]. Los sistemas recomendadores han sido y están siendo aplicados en diversidad de campos, así destacan las aplicaciones efectuadas en el ámbito de turismo como es el Sistema consultor de viajes [5], en finanzas Warren [2], en restauración IRES [3], o en comercio electrónico el CDNow y Amazon [8]. En un sitio *web* los agentes recomendadores adquieren información de los usuarios por ejemplo rastreando las páginas que visita y el tiempo que permanecen en ella y usan estos datos para analizar la relación entre sus intereses y cuales productos actualmente ellos compran. Estos datos dan una idea real de los intereses y comportamiento de los usuarios [8].

Para hacer eficiente el proceso de conseguir información de los consumidores potenciales, analizarla y tomar decisiones basados en ella, se definieron un conjunto de medidas que ayudan a tener una idea a priori de la información útil contenida en las fuentes de información antes de su utilización. Estas medidas junto con el *trust* de las recomendaciones pasadas de fuentes similares pueden predecir el resultado de recomendaciones futuras.

El resto del artículo esta organizado de la siguiente manera, en la **Sección 2** se presentan las medidas de las fuentes de información. La **tercera sección** define el valor esperado de futuras recomendaciones. La **Sección 4** presenta un caso de estudio realizado en el dominio de ventas de bienes de gran consumo. Finalmente en la **Sección 5** se presentan las conclusiones.

## 2. Coleccionando datos de los consumidores

Una fuente de información son las bases de datos que permiten obtener los datos de los clientes actuales y o potenciales de las empresas. Por lo tanto estas fuentes contienen la información que permite ofrecer a los clientes aquellos productos o servicios que puedan resultar de su interés. Algunos de los datos utilizados son: datos sociodemográficos, profesionales, datos de consumo que son los que se refieren a las compras realizadas, o a la utilización de los servicios, etc. Para que la información sea rentable en la empresa es necesario que esté actualizada, que sean lo suficientemente representativa, que dispongan del suficiente universo de muestras para su explotación, y que se pueda conocer el estado de la información contenida en ella.

En este trabajo se definieron un conjunto de medidas que permiten conocer a priori la información contenida en las fuentes que servirá para quienes tomen las decisiones de marketing o a en un sistema recomendador. Estas características permiten tener:

- Una representación de la información contenida en cada fuentes y,
- Un criterio para comparar y establecer similaridad entre las fuentes.

Cuando se procede a buscar la información, tanto interna como externamente, no toda la información es útil, sería bueno contar con una clasificación de las fuentes de información que permita conocer en forma abstracta la información contenida en ellas. Para no realizar una búsqueda exhaustiva se tendría que tener catalogada o indexada la montaña de información para que se pueda encontrar la necesaria con mayor facilidad. Las medidas propias de las fuentes permitirían tener una idea a priori de la información contenida en ellas como lo realiza un catálogo de información. Con estas características se podría obtener detalles útiles para la segmentación de grupos de consumidores, cantidad de posibles clientes y estado de la información.

La Figura 1 muestra las medidas definidas para este propósito [1]:

Características	
<b>Compleitud:</b> Número de usuarios de una fuente encontrados en otra fuente	$Compleitud\ d = \frac{\sum(A \cap B)}{\sum A}$ <p><math>(A \cap B)</math> = Usuarios existentes en ambas fuentes  <math>A, B</math> = Usuarios de una fuente</p>
<b>Diversidad:</b> Número de grupos de usuarios. Permite agrupar a los usuarios de acuerdo a un criterio dado	$H = -\sum (p_i * \ln p_i) \quad p_i = \frac{n_i}{N}$ <p><math>n_i</math> = Cantidad de usuarios en el grupo i  <math>N</math> = Total de usuarios en la fuente</p>
<b>Ontología:</b> Representación semántica de la información contenida en las fuentes	$Ontología = \frac{a}{A}$ <p><math>a</math> = Cantidad de atributos relevantes encontrados en la fuente.  <math>A</math> = Total de atributos relevantes para la recomendación.</p>
<b>Frecuencia:</b> Frecuencia de las interacciones de los usuarios con la fuente de información	$Frecuencia = \frac{\sum w_i * f_i}{N}$ <p><math>f_i</math> = Cantidad de usuarios que compraron en un periodo de tiempo I  <math>w_i</math> = Peso otorgado al periodo de tiempo  <math>N</math> = Cantidad de usuarios de la fuente</p>
<b>Actualización:</b> Actualización de la información contenida en las fuentes	$Actualización = \frac{\sum w_i * c_i}{N}$ <p><math>C_i</math> = Cantidad de usuarios que interaccionaron en un determinado radio de frecuencia.  <math>w_i</math> = Peso del radio de frecuencia de compra  <math>N</math> = Cantidad de usuarios de la fuente.</p>

Figura 1: Características Intrínsecas de las fuentes de información

### 3. Valor esperado de la recomendación

En una recomendación el resultado puede ser incierto pero basándose en la experiencia de recomendaciones pasadas y las características de las fuentes de información utilizadas para realizarlas se puede asignar una valoración a priori a partir de los porcentajes de veces que fue exitosa o no, estas proporciones son representadas por probabilidades.

La eficacia o éxito de una recomendación es medida por su precisión de acuerdo a la siguiente ecuación [7].

$$Precision = \frac{NR}{n} \quad (1)$$

Donde  $NR$  es la cantidad de productos recomendados que fueron comprados por los clientes y  $n$  es el número de productos recomendados

Utilizando esta medida se puede calcular el valor esperado de la recomendación como la esperanza matemática:

$$E = \sum x_i p(X = x_i) \quad (2)$$

Donde  $i = \{exitosa, no\_exitosa\}$  y  $x_i$  es el peso otorgado cuando la recomendación es exitosa o no exitosa. El éxito de las recomendaciones es medida por la precisión, si las medidas son más altas que un valor determinado de mejora significa que la recomendación fue exitosa.

$$p(X = Exitosa) = \frac{E}{N} \quad (3)$$

$$p(X = No\_Exitosa) = \frac{NE}{N} \quad (4)$$

Donde  $E$  es el número de recomendaciones exitosa,  $NE$  es el número de recomendaciones no exitosas y  $N$  es el total de recomendaciones.

### 3.1 Ejemplo

Al resultado de la recomendación se le otorgará el valor de exitosa o no\_exitosa de acuerdo a la siguiente regla:

*Si ( precisión > 0,6 )  
Entonces resultado= Exitosa  
Si no resultado= No\_exitosa*

Después de aplicar esta regla se obtiene una tabla como la siguiente (ver Tabla 1):

<b>Recomendación</b>	<b>Resultado</b>
1	Exitosa
2	No_exitosa
3	No_exitosa
4	Exitosa
5	Exitosa
6	Exitosa
7	No_exitosa
8	Exitosa
9	No_exitosa
10	Exitosa
11	Exitosa
12	Exitosa

Tabla 1: Evaluación de las recomendaciones

Dado  $x_{exitosa} = 0.7$  y  $x_{no\_exitosa} = 0.3$

El valor esperado de la recomendación es:  $E = (0.7 * 8/12) + (- 0.3 * 4/12) = 0.327$

### 3.2 Valor esperado como una medida de *trust* de las recomendaciones

El valor esperado puede ser calculado como una medida del *trust* como lo define Patel en [4].

$$Trust(a_1, a_2) = E[B_{a_2} / \alpha, \beta] \quad (5)$$

Donde:

$$E[B_{a_2} / \alpha, \beta] = \frac{\alpha}{\alpha + \beta} \quad (6)$$

$$\alpha = m_{a_1, a_2}^{lr} + 1 \quad (7)$$

$$\beta = n_{a_1, a_2}^{lr} + 1 \quad (8)$$

Esta medida de *trust* es basada en las interacciones de dos agentes en un SMA.

Donde  $m_{a_1, a_2}^{lr}$  es el número de interacciones exitosas y  $n_{a_1, a_2}^{lr}$  es el número de interacciones no exitosas.

Al emplear esta medida de *trust* en el caso de las recomendaciones (R),  $m_{a_1, a_2}^{lr}$  sería el número de recomendaciones exitosas y  $n_{a_1, a_2}^{lr}$  el número de recomendaciones no exitosas .

De esta manera se puede plantear el valor esperado del resultado de una recomendación como una medida de *trust*:

$$E(R) \equiv \mathbf{Trust}(R) \quad (9)$$

### 3.3 Valor esperado de la recomendación teniendo en cuenta las características de las fuentes de información.

Si se desea utilizar una nueva fuente de información se puede conocer a priori el valor de eficacia de las recomendaciones haciendo uso de sus características y el *trust* de las recomendaciones pasadas de una fuente similar a ella, esto es:

#### Premisas:

- 1) Dadas dos fuentes S1 y S2 representada por sus características:

$$\begin{array}{ll} S1 = \{Ci\} & Ci = \{\text{Ontología}, \dots, \text{frecuencia}\} \\ S2 = \{Ni\} & Ni = \{\text{Ontología}, \dots, \text{frecuencia}\} \end{array}$$

2) Las recomendaciones realizadas con información de las dos fuentes

$$R(S1+S2)$$

3) Si existe una nueva fuente de información S3

$$S3 = \{Mi\} \quad Mi = \{\text{Ontología}, \dots, \text{frecuencia}\}$$

Y  $S3=S2$ , los valores de las características son los mismos:

$$\{Mi\} = \{Ni\}$$

### **Hipótesis:**

Se espera que  $E(R(S1+S2)) = E(R(S1+S3))$

Como  $E(R(S1+S2)) \equiv \text{Trust}(R(S1+S2))$

Entonces  $E(R(S1+S3)) \equiv \text{Trust}(R(S1+S2))$

De esta manera se puede evaluar *a priori* y sin haber utilizado antes esta fuente de información S3 cual sería el valor de eficacia de un sistema recomendador a partir del *trust* de fuentes similares.

## **4. Caso de estudio**

En el caso de estudio se utilizaron tres bases de datos en el sector de venta de bienes de gran consumo (supermercado). Una base de datos la S1 tiene información de las compras realizadas por Internet (online), las otras con información de las compras realizadas en el supermercado.

La información contenida en ellas es de 1200 clientes y las compras realizadas durante el periodo 2001-2002. Las bases de datos tienen clientes comunes que tienen el mismo identificador.

Para cada una de las bases de datos se calcularon sus medidas intrínsecas como lo muestra la siguiente tabla:

Características	Fuente 1 (S1)	Fuente 2 (S2)	Fuente 3 (S3)
Ontología	0.80	0.50	0.50
Diversidad			
Z(Zona)	0.13	0.11	0.11
F(Familia)	0.66	0.67	0.67
S(Sexo)	0.20	0.20	0.20
Compleitud	0.10	0.60	0.60
Frecuencia	0.23	0.40	0.40
Actualización	0.25	0.40	0.40

Tabla 2: Medidas intrínsecas de las fuentes de información en el sector de venta de bienes de gran consumo (supermercado)

Como se observa la base de datos más completa, es la S2 y S3 que tienen las mismas características, tiene más cantidad de clientes, la más actualizada y la que tiene más cantidad de transacciones de los clientes .

Para realizar las recomendaciones se utilizó la información de la fuente S1 y S2. Se utilizó el método TF-IDF (*term frequency times inverse document frequency*) [6] para establecer la relevancia de cada producto para los usuarios a partir de compras previas.

$$W_{ik} = T_{ik} * \log_2\left(\frac{N}{n_k}\right) \quad (10)$$

Donde  $T_{ik}$  es la frecuencia del producto  $k$  en las compras del cliente  $i$ ,  $n_k$  es la cantidad de clientes que compraron el producto  $k$  y  $N$  es la cantidad total de clientes.

Para todos los clientes de las fuentes S1 y S2 se realizaron las recomendaciones por un periodo de 6 semanas. Las recomendaciones realizadas de una semana pueden ser observadas en la Figura 2:

Codi_client	RECOMENDADO
Cliente336Z2F+4OFF	ESPINACA 250 G I FREDECA
Cliente336Z2F+4OFF	SIN RETORNO 1,25 L FONTER
Cliente336Z2F+4OFF	NORMAL 2 L COCA COLA
Cliente336Z2F+4OFF	SIN RETORNO 1,25 L FONTER
Cliente336Z2F+4OFF	ESPINACA 250 G I FREDECA
Cliente336Z2F+4OFF	12 ROLLO HIGIENI CAPRABO
Cliente336Z2F+4OFF	NORMAL 2 L COCA COLA
Cliente636Z2F3OFF	CERVEZA LATA 33CL CAPRABO
Cliente636Z2F3OFF	ACEIT. RELLENAS 150 G CAPRABO
Cliente1067Z2F3OFF	NORMAL 2 L COCA COLA
Cliente1067Z2F3OFF	12 ROLLO HIGIENI CAPRABO
Cliente1067Z2F3OFF	LATA LIMON 33CL SCHWEPPE
Cliente1067Z2F3OFF	BRIK DES. CALCIO 1 L PULEVA
Cliente1067Z2F3OFF	GEL BA#O AVENA 750ML AVENA KINE
▶ Cliente1067Z2F3OFF	MAIZ DULCE X3 450 G CAPRABO
Cliente2765Z2F2OFF	LIMPIAHO. ANTIB. 1500ML TENN
Cliente2765Z2F2OFF	LATA TONICA LIGH 33CL SCHWEPPE
Cliente2765Z2F2OFF	ACEITE V. ARBEQUI 75CL BORGES
Cliente2912Z2F2OFF	GEL DERM 750ML SANEX 750
Cliente2912Z2F2OFF	ACEITE EX. VIR 04 1L HOJIBLANCA
Cliente2912Z2F2OFF	BOT. ENTERA 1,5 L ASTURIANA
Cliente2912Z2F2OFF	BOT. ENTERA 1,5 L ASTURIANA
Cliente3102Z2F4OFF	PRINCIPE ESTREL 150 G BEUKELAER
Cliente3715Z2F4OFF	ANTICAL 500ML VIAKAL
Cliente3715Z2F4OFF	12 ROLLO HIGIENI CAPRABO
Cliente3715Z2F4OFF	PATATA ESP. FREIR 3KG
Cliente3715Z2F4OFF	AZUCAR FINO 1000 G AZUCARERA
Cliente3715Z2F4OFF	VINAGRE BLANCO 1 L CAPRABO
Cliente3715Z2F4OFF	BRIK SEMIDESNAT. 1 L CAPRABO
Cliente3966Z2F2OFF	COMBI SEMIDESNAT 1 L ASTURIANA
Cliente3966Z2F2OFF	PIZZA ATUN/BACON 425 G TARRADELLA
Cliente3966Z2F2OFF	JAMON LON. EXTRAFINAS 1 ESPU#A
Cliente3966Z2F2OFF	AZUCAR FINO 1000 G AZUCARERA
Cliente3981Z2F2OFF	PLATANOS FLOW-PACK

Figura 2: Recomendaciones realizadas durante una semana con información de los clientes de la fuente S1 y S2

Al evaluar las recomendaciones durante cada semana se obtuvo los siguientes resultados:

Recomendaciones	Eficacia (Precisión)	Valoración
R1	0.5	Exitosa
R2	0.6	Exitosa
R3	0.2	No_Exitosa
R4	0.3	No_Exitosa
R5	0.8	Exitosa
R6	0.6	Exitosa

Tabla 3: Evaluación de las recomendaciones realizadas con S1 y S2

Antes de realizar las recomendaciones con la fuente S3 se puede saber el valor de eficacia de la recomendación a partir del *trust* de las recomendaciones de las fuentes S1 y S2 por que la fuente S3 tiene es similar a la fuente S2.

$$E(R(S1+S3)) \equiv \mathbf{Trust}(R(S1+S2))$$

Por que  $S2=S3$ .

Aplicando ecuaciones (5) y (6)

$$\mathit{Trust}(R(S1 + S2)) = \frac{\alpha}{\alpha + \beta} = \frac{5}{5 + 3} = 0,62$$

Por lo que se espera que el valor de eficacia de las recomendaciones con información de las fuentes S3 sea igual a 0,62.

Para comprobar esta hipótesis se realizaron las recomendaciones y se obtuvo un valor de eficacia de:

$$\mathbf{Precisión R(S1+S3)= 0,58}$$

El resultado obtenido no fue exactamente 0,62 pero es un resultado que se puede considerar bastante similar lo que demuestra que el *trust* de las recomendaciones de fuentes de información similares a la que se desea utilizar tiene dan una idea a priori de la eficacia de las recomendaciones.

## 6. Conclusiones

Las principales aportaciones de este trabajo es la elaboración de un conjunto de métricas propias de las fuentes de información que pueden ser usadas en un sistema de recomendación como criterio de selección de las fuentes antes de su utilización a partir de la medida de *trust* de fuentes similares.

De los resultados obtenidos se observa que el *trust* de las fuentes de información similares dan una idea a priori de la eficacia de recomendaciones futuras. Los valores de las medidas no significan que el contenido de las fuentes sea el mismo, por ejemplo que dos fuentes de información tengan el mismo valor de ontología, no significa que posean los mismos atributos iguales necesarios para las recomendaciones.

## Referencias

- [1] Aciar Silvana, López Herrera, Josefina y De la Rosa Josep Lluís; SMA para la búsqueda inteligente de información para recomendar, I Congreso Español de Informática, Cedi 2005, Simposio de Inteligencia Computacional, SICO'2005 (IEEE Computational Intelligence Society, SC) Septiembre de 2005, Granada, España.
- [2] Jennings, N.R., and Wooldridge, M.J., (eds) Agent Technology: Foundations, Applications, and Markets, ISBN:3-540-63591-2 Springer, 1998.
- [3] M. Montaner, B. López, E. del Acebo, S. Aciar, I. Cuevas IRES: On the Integration of Restaurant Services AgentCities Agent Technology Competition. Barcelona, Spain, Feb 2003. Available at: <http://eia.udg.es/ar/ires.pdf>
- [4] Patel, J., Teacy, W. T. L., Jennings, N. R. and Luck, M. (2005) A Probabilistic Trust Model for Handling Inaccurate Reputation Sources. In Proceedings of Third International Conference on Trust Management (in press), Rocquencourt, France. Herrmann, P., Issarny, V. and Shiu, S., Eds.
- [5] Ricci, F. Travel Recommender systems. IEEE Intelligent Systems, 7(6), pp. 55-57, 2002.
- [6] G. Salton and M.J.McGill: Introduction to Modern Information Retrieval. McGraw-Hill Publishing Company, New York, NY, 1983.
- [7] G. Salton. Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer. Addison-Wesley, 1989
- [8] Schafer, J. B., Konstan, J. and Riedl, J.: Recommender Systems in E-Commerce. In: EC '99: Proceedings of the First ACM Conference on Electronic Commerce, Denver, CO, pp. 158-166. 1999