

Obteniendo valoraciones de ítems desde opiniones de usuarios para resolver el problema de “Cold Start” en sistemas recomendadores

Silvana Aciar¹, Gabriela Aciar¹, Nestor Duque²

¹ Instituto de Informática
Universidad Nacional de San Juan, Argentina
Av. Central y Meglioli
Rivadavia, San Juan, Argentina - 5400
¹ saciar@iinfo.unsj.edu.ar, gaby_aciar@yahoo.com.ar
² Universidad Nacional de Colombia – Sede Manizales
Manizales - Colombia
ndduqueme@unal.edu.co

Abstract. Un problema bien conocido de los sistemas recomendadores es el problema denominado “Cold Start”, que es causado por la falta de información de usuarios o productos/servicios. Un sistema de recomendación sólo puede producir buenas recomendaciones después de haber acumulado suficientes datos. El problema se vuelve aún más difícil cuando el sistema de recomendación trata de hacer frente a nuevos productos o los productos no han sido valorados por los consumidores. En este trabajo se aborda este problema obteniendo valoraciones de los productos/servicios desde los comentarios escritos por usuarios en foros. Un caso de estudio de recomendación de objetos de aprendizajes es presentado.

Keywords: Sistemas recomendadores, personalización, minería de texto, e-learning.

1. Introducción

Los sistemas de recomendación son aplicaciones que ayudan al usuario a elegir un producto o servicio. Son empleados con gran éxito en el comercio electrónico, ofreciendo productos o servicios personalizados a los clientes en base a sus preferencias e intereses. La mayoría de los sistemas de recomendación existentes utilizan los tres métodos más populares, filtrado colaborativo, filtrado basado en contenido y el filtrado híbrido, que combina los dos métodos anteriores [1]. Métodos de recomendación basados en el filtrado colaborativo sugieren productos o servicios en base a las preferencias de otros usuarios con perfiles similares. Por el contrario, los métodos basados en contenido utilizan la información sobre el objeto en sí para realizar las sugerencias.

El gran problema del filtrado colaborativo, que produce recomendaciones mediante la comparación de las selecciones anteriores de un cliente con otros clientes que han hecho selecciones similares, es el denominado “Cold Start Problem” [2][3]. Al inicio de la utilización de estos sistemas, hay poca o ninguna información inicial del usuario disponible para hacer recomendaciones. Para que estos sistemas puedan dar recomendaciones más precisas es necesario que los consumidores valoren una gran cantidad de productos y así obtener sus gustos y preferencias. Existen dominios de aplicación de los sistemas recomendadores tales como turismo, educación, entretenimiento y otros donde los usuarios o consumidores no están dispuestos a valorar una gran cantidad de ítems. En estos dominios existen cada vez más evidencias de que con el surgimiento de las redes

sociales, foros y blogs, los usuarios prefieren expresar sus preferencias, opiniones e intereses a través de estas herramientas. Hoy en día, los foros de opiniones de productos y grupos de discusión son formas populares para que los consumidores intercambien sus experiencias con un producto [4] [5]. Hay una creciente evidencia de que este tipo de foros influyen las decisiones de compra de los consumidores [6] [7].

Por lo tanto, adquiriendo información de estas fuentes se puede solucionar el problema de la falta de información inicial. En investigaciones anteriores se ha propuesto un mecanismo integral que formaliza el proceso de selección y recuperación de opiniones, como así también la estructuración de esas opiniones en una ontología y la obtención del producto mejor valorado se recomienda [8].

En este artículo se aplica ese mecanismo integral para estimar valoraciones de productos de las opiniones escritas en formato texto y utilizar esas valoraciones en un sistema recomendador aplicando el filtrado colaborativo. Resultados preliminares muestran que las precisiones de las recomendaciones aumentan si se tiene más valoraciones de los usuarios resolviendo el problema del “Cold Start”.

El resto del trabajo se organiza de la siguiente manera: la sección 2 ofrece una breve introducción al proceso de obtener las valoraciones de los productos desde comentarios de usuarios en los foros. La sección 3 presenta el problema de “Cold Start”. Un caso de estudio y resultados experimentales en el dominio de e-learning son presentados en la sección 4 y 5 respectivamente. Finalmente, la sección 5 concluye el documento.

2. Adquiriendo información de las opiniones de usuarios

Aunque hay gran cantidad de opiniones de los consumidores publicadas en diversas fuentes disponibles en Internet, la información no ha sido ampliamente utilizada por los sistemas de recomendación, debido a las dificultades de procesamiento de datos en formato texto. El sistema de recomendación propuesto en [8] contiene tres componentes:

- Representación de la información en un formato adecuado que puede ser utilizado por el sistema (generación de una ontología).
- Cálculo de la valoración del producto de las opiniones.
- Selección del producto más valorado para su recomendación.

Este sistema de recomendación se muestra en la Figura 1.

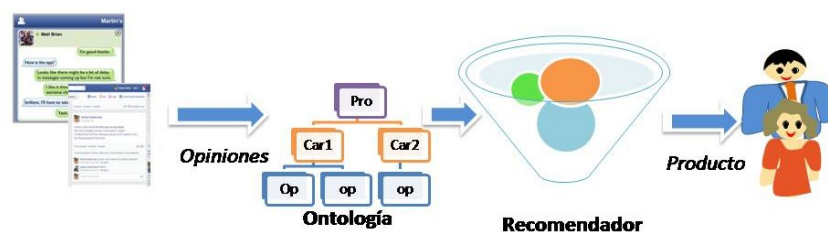


Fig. 1. Sistema de recomendación en base a información obtenida desde las opiniones de usuarios [8]

Los comentarios de los usuarios son analizados, se seleccionan los más relevantes, se clasifican y se asignan a la ontología. La ontología contiene dos partes principales: Las características del producto y las opiniones de los usuarios para cada una de las características. Estas opiniones se clasifican en Positivas o Negativas. Un conjunto de medidas ha sido definido para obtener el valor global del producto: Valor de las características, en base a la información de la ontología se obtiene un valor numérico de cada característica del objeto para un usuario. Valor del producto para un usuario, resume la valoración que le da un usuario al producto en base a las valoraciones de todas las características en un solo valor. Valor total del producto para todos los usuarios, es un valor numérico obtenido a partir de la valoración del mismo producto por varios usuarios. Más detalle del cálculo de estas medidas se puede obtener en [8].

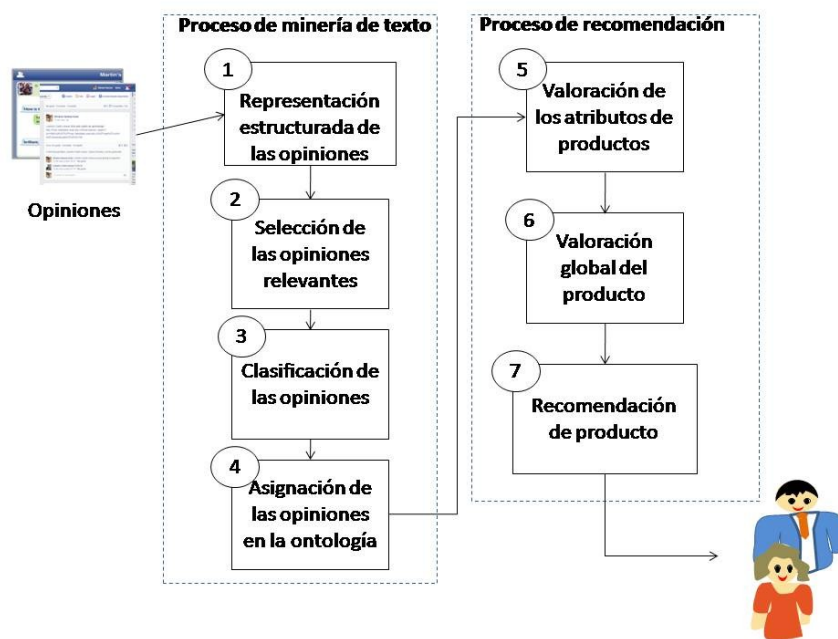


Fig. 2. Proceso de obtención de valoraciones de los productos desde opiniones escritas en foros

La recomendación se hace en base a los comentarios que se resumen en una sola medida, Valoración Global del Producto (Figura 2).

En este artículo se emplea la medida de Valoración del Producto para un Usuario para así obtener la valoración de usuarios sobre productos que no fueron valorados explícitamente y poder realizar el filtrado colaborativo.

3. El problema de “Cold Start”

El problema de Cold Start se puede clasificar en dos tipos: el problema de un nuevo sistema y el problema de nuevos ítems/usuario. El problema del nuevo sistema es donde hay poca o ninguna información inicial para que el sistema realice recomendación, este problema se puede eliminar en gran medida gracias a la gran cantidad de comentarios disponibles en Internet hoy en día. Sin embargo,

los comentarios proporcionados por los consumidores no son de todos los productos existentes en un sitio. El problema de recomendar productos sin valoraciones, en este caso, es similar al problema un nuevo producto / usuario, donde el sistema ha estado funcionando durante un tiempo y un nuevo producto / usuario aparece, el sistema no dispone de información para realizar la recomendación. La mayoría de los sistemas de recomendación funcionan mal en esta situación [2] [3] .

A continuación se presenta un caso de estudio donde, por medio de la obtención de valoraciones desde las opiniones escritas en la web, se puede solucionar el problema de Cold Start, aumentando la precisión de las recomendaciones.

4. Caso de estudio

Se seleccionó el dominio de E-learning para realizar recomendaciones de objetos de aprendizaje a los estudiantes. Para emplear los métodos de recomendación como el Filtrado Colaborativo, se necesita que los objetos sean valorados por los usuarios. Es difícil obtener valoración de los objetos de forma explícita por los docentes o estudiantes que utilizaron los objetos previamente. Una fuente muy relevante para obtener estas valoraciones son los foros en las plataformas virtuales que disponen las instituciones educativas. En esta sección, se presenta la extensión del caso de estudio presentado en [8]. Se utilizó la plataforma virtual de aprendizaje Moodle. De dicha plataforma se han coleccionado los comentarios en los foros donde se involucraba la discusión de objetos de aprendizajes en diferentes cursos.

La extensión consiste en la presentación de resultados obtenidos al realizar recomendaciones de objetos de aprendizajes utilizando el método del filtrado colaborativo. El objetivo de estos experimentos es evidenciar el hecho que se pueden obtener valoración de los objetos de aprendizajes desde las opiniones de los usuarios para aumentar las precisiones de las recomendaciones. En el trabajo anterior, solo se recomendaban los objetos mejor valorados por los usuarios, a cada usuario se presentaba los mismos objetos. Aquí las recomendaciones son realizadas de forma personalizadas.

Se ha utilizado la plataforma virtual de aprendizaje Moodle, donde se ha puesto a disposición de 32 usuarios de la plataforma 8 objetos de aprendizajes para que ellos opinen en diferentes foros. En la figura 3 se presenta opiniones sobre un objeto de aprendizaje obtenida como resultado de la selección de las opiniones relevantes.

Cabe destacar que se ha elaborado una lista conteniendo las palabras usualmente escritas de manera no formal por los estudiantes, por ejemplo, los estudiantes suelen escribir X en lugar de la palabra “por” o Q por “que”. Esta lista ha sido elaborada manualmente.

Objeto: Ecuaciones lineales

Usuario: AnaF, 26 abril 2014
Es un buen recurso. Pero no lo puedo instalar en mi computadora. Me sale un error cuando lo quiero abrir y no encuentro alguna especificación de requerimientos de programas.

Usuario: RafaProfe, 29 abril 2014
Que buen video, lo que necesitaba. Las explicaciones claras, los ejemplos bien aplicados, una buena didáctica. Y bien dirigido para niños pequeños a los cuales está dirigido. El unico problema es el audio. No se puede escuchar aunque esté subtitulado.

Usuario: Josefina_2009 , 3 mayo 2014
Soy invidente, aunque el video dice que posee audio, no se escucha

Fig. 3. Opiniones sobre un objeto de aprendizaje obtenidas como resultado del pre-procesamiento [8].

4.1. Asignación de información de los comentarios en la ontología

Se ha utilizado la ontología definida en [8] para estructurar la información procedente de las opiniones. Para que la información de los comentarios se pueda asignar a la estructura de la ontología se realizó la clasificación de las sentencias de las opiniones y luego la identificación de conceptos.

Para la clasificación de sentencias se implementó una rutina que identifica en cada sentencia las palabras negativas como por ejemplo “no, problema, etc.” o las palabras positivas como por ejemplo “sí, bueno, excelente, etc” pudiendo clasificar de acuerdo a esto una sentencia en “Bueno” o “Malo” como se presenta en la Figura 4.

```
Sentencia 1:      Es un buen recurso.
Clasificación: Bueno
-----

Sentencia 2:      Pero no lo puedo instalar en mi
computadora.
Clasificación: Malo
-----

Sentencia 3:      Me sale un error cuando lo quiero abrir
y no encuentro alguna especificación de requerimientos de
programas.
Clasificación: Malo
```

Fig. 4. Clasificación de sentencias en las categorías “Bueno” y “Malo”.

4.2. Identificación de conceptos de la ontología en las sentencias

Para cada sentencia que se encuentra en las categorías "Bueno" o "Malo", la función de mapeo se realiza buscando las palabras relacionadas con los conceptos de la ontología en cada una de las sentencias. Por ejemplo la palabra “programas” se relaciona con el concepto “Recurso Tecnológico Requerido” de la ontología (Figura 5).

```
Sentencia 3:      Me sale un error cuando lo quiero abrir
y no encuentro alguna especificación de requerimientos de
programas.

Clasificación: Malo

Concepto ontología: Recurso Tecnológico Requerido
(palabra relacionada: Programas)
```

Fig. 5. Identificación de conceptos de la ontología en las sentencias de una opinión.

Un ejemplo de la ontología obtenida como resultado del procesamiento de las opiniones de usuarios se puede observar en la figura 6.

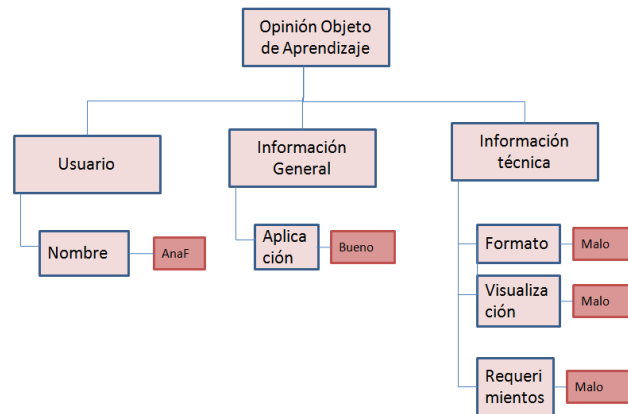


Fig. 6. Ontología obtenida conteniendo información de las opiniones de un usuario.

5. Resultados experimentales

Dos experimentos fueron realizados utilizando el método del filtrado colaborativo. La métrica utilizada ha sido la métrica del Coseno (Ecuación 1).

$$\cos(u_i, u_j) = \frac{\sum_{k=1}^m v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^m v_{ik}^2 \sum_{k=1}^m v_{jk}^2}} \quad (1)$$

Donde u_i, u_j son representados como vectores. v_{ik} es la valoración del usuario i para el producto k . v_{jk} es la valoración del usuario j para el producto k . La obtención de las valoraciones en ambos experimentos fue diferente.

Experimento 1: Recomendación de los objetos sin tener en cuenta información desde las opiniones expresadas en forma textual.

Las valoraciones de los objetos por los usuarios han sido obtenidas mediante valoración explícita por los estudiantes.

Al iniciar los foros en los cursos utilizados en este experimento se les indicó a los usuarios que valoren los objetos de aprendizajes que se vayan presentando en el curso, para ello se les dio una escala de 1 al 5 para su valoración.

El sistema recomendador utiliza esta información para realizar las recomendaciones utilizando la ecuación 1.

El sistema recomendador presentó a cada usuario una lista de tres objetos de aprendizajes.

Experimento 2: Recomendación de los objetos teniendo en cuenta información desde las opiniones expresadas en forma textual.

Las valoraciones de los objetos por los usuarios han sido obtenidas desde los comentarios que los estudiantes han escrito en los foros acerca de los objetos. En cada curso se les informó a los estudiantes que podían hacer comentarios en los

foros acerca de los objetos de aprendizajes con los cuales se trabajaba. Esa información ha sido la utilizada por el método descrito en la Sección 2 y estructurada en la ontología.

El sistema recomendador utiliza esta información para realizar las recomendaciones utilizando la ecuación 1.

El sistema recomendador presentó a cada usuario una lista de tres objetos de aprendizajes.

Para evaluar las recomendaciones realizadas en ambos experimentos, se utilizó la medida de precisión obtenida a partir de la ecuación 2.

La evaluación se realizó teniendo en cuenta las recomendaciones exitosas sobre todas las recomendaciones realizadas. Para obtener la retroalimentación de los usuarios acerca de las recomendaciones realizadas, se les pidió que valoraran las recomendaciones realizadas.

$$\text{Precisión} = \frac{NR}{N} \quad (2)$$

Donde NR es el número de recomendaciones exitosas y N el número total de recomendaciones realizadas. Esta medida se ha obtenido por cada sesión realizada y para ambos experimentos. Los resultados se pueden observar en la figura 7. Como se puede observar las recomendaciones realizadas en base a las valoraciones obtenidas de las opiniones de los usuarios son más altas que las realizadas con información de valoración de los usuarios en la forma tradicional (utilizando una escala). Analizando las causas de la diferencia entre los dos experimentos, se observó que muchas similitudes entre usuarios no se podían realizar en el experimento 1 debido a que muchos objetos no fueron valorados por los usuarios. Estos mismos objetos, en el experimento 2 si fueron valorados por los usuarios mediante comentarios en los foros, los usuarios fueron más propensos a valorar objetos mediante los comentarios.

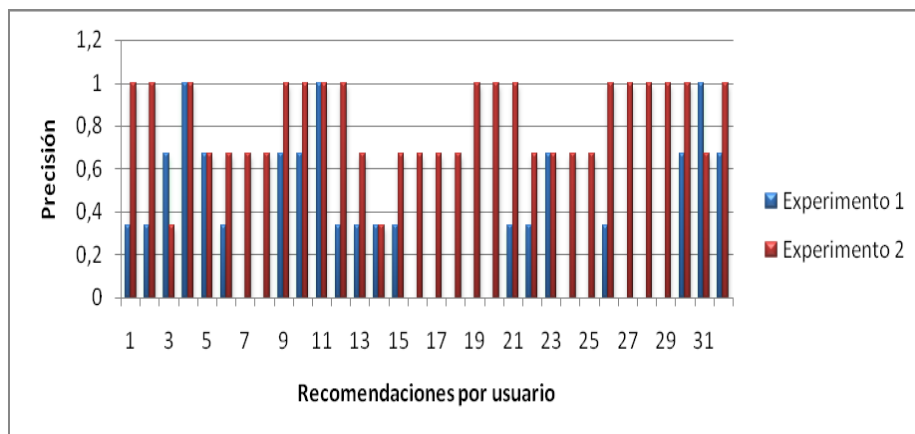


Fig. 7. Precisión de las recomendaciones en ambos experimentos

6. Conclusiones

En este trabajo se aborda el problema de “Cold Start” en los sistemas de recomendación en dominios donde los productos son pocos valorados. Existen dominios de aplicación de los sistemas recomendadores tales como turismo, educación, entretenimiento y otros donde los usuarios o consumidores no están dispuestos a valorar una gran cantidad de ítems. La solución proporcionada fue considerar las opiniones de los usuarios que realizan en los foros. El caso de estudio demuestra que al obtener más valoraciones se obtienen recomendaciones más precisas. Una forma de obtener más valoraciones es analizando lo que los usuarios pueden expresar a través de comentarios en redes sociales, foros y blogs. Los usuarios prefieren expresar sus preferencias, opiniones e intereses a través de estas herramientas.

Agradecimientos

Este trabajo está enmarcado en la Red Temática 513RT0481 “Red Iberoamericana de apoyo a los procesos de Enseñanza-Aprendizaje de competencias profesionales a través de entornos ubicuos y colaborativos” financiada por el Programa CYTED, “Desarrollo De Herramientas Tecnológicas De Soporte A La Educación Virtual” financiado por CICITCA-UNSJ y “E-OASIS Personalizando información turística en Internet para promocionar el patrimonio cultural” PRH PICT- 00035.

Referencias

1. Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4):331-370, 2002.
2. Blerina Lika, Kostas Kolomvatsos, Stathes Hadjiefthymiades, Facing the cold start problem in recommender systems, *Expert Systems with Applications*, Volume 41, Issue 4, Part 2, March 2014, Pages 2065-2073, ISSN 0957-4174.
3. Guibing Guo , Jie Zhang , Daniel Thalmann, Merging trust in collaborative filtering to alleviate data sparsity and cold start, *Knowledge-Based Systems*, 57, p.57-68, February, 2014
4. J. Tang,X. Hu,H. Liu Social recommendation: a review *Social Network Analysis and Mining* 3 1113-1133 2013.
5. Fengkun Liu , Hong Joo Lee, Use of social network information to enhance collaborative filtering performance, *Expert Systems with Applications: An International Journal*, v.37 n.7, p.4772-4778, July, 2010
6. V. Zanardi and L. Capra, "A Scalable Tag-Based Recommender System for New Users of the Social Web, " in *Database and Expert Systems Applications*. vol. 6860, A. Hameurlain, S. Liddle, K.-D. Schewe, and X. Zhou, Eds., ed: Springer Berlin Heidelberg, 2011, pp. 542-557
7. Xujuan Zhou, Yue Xu, Yuefeng Li, Audun Josang, and Clive Cox. 2012. The state-of-the-art in personalized recommender systems for social networking. *Artif. Intell. Rev.* 37, 2 (February 2012)
8. Aciar, S. Valoración de objetos de aprendizajes desde comentarios escritos por usuarios en la web. *XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014)*. 2014.