Creado el 13/01/2014
Última versión: 13/01/2014
Arturo Montejo Ráez, Universidad de Jaén
Comunidad moocTLH
Esta obra está bajo una licencia de Creative Commons Reconocimiento-CompartirIgual 3.0 Unported.

Retos y oportunidades de las TLH/PANORAMA ACTUAL

SISTEMAS DE RECOMENDACIÓN

A todos nos suena el término "fidelización" o alguna vez hemos buscado la opinión de otros a la hora de comprar un producto, sea por Internet, sea en una tienda física. Veamos las líneas generales de cómo se construye un sistema de este tipo aprovechando las TLH.

Transcripción del vídeo

Cualquiera de nosotros alguna vez ha comprado un producto siguiendo los consejos o sugerencias del tendero de nuestro barrio. Él sabe perfectamente que un cliente satisfecho es un cliente que probablemente volverá a comprar en el establecimiento, por lo que una buena recomendación es una forma de fidelizarlo. Pero, claro está, la sugerencia debe satisfacer las necesidades del cliente y eso implica conocer bien sus gustos y preferencias. ¿Qué estrategia sigue nuestro tendedero? Pues saber cuál es el tipo de producto por el que solemos optar, proponer nuevos productos relacionados con nuestro perfil de compra o sugerir aquello que fue del gusto de otro comprador con tendencias similares.

Con la llegada del comercio electrónico en general y la compra por Internet en particular, han surgido a la par los sistemas de recomendación automáticos. Este tipo de sistemas es el que tratamos en esta lección, mostrando su funcionamiento y algoritmos principales. Veremos cómo el procesamiento del lenguaje natural puede llegar a ser un factor importante en los mismos. Los sistemas de recomendación son de gran interés para la industria y es mucho el esfuerzo invertido en lograr la satisfacción de los clientes a través de los mismos. Tanto es así que, en el año 2007, Netflix, una poderosa empresa de alquiler de películas por Internet, convocó una competición en la que ofreció un millón de dólares a aquella propuesta que mejorara por encima del 10% a Cinematch, su algoritmo de predicción de las puntuaciones de los clientes a películas. Llevó tres años que un grupo de investigadores lograran el 10.06% de mejora, gracias a un algoritmo de filtrado colaborativo, uno de los comentaremos en breve. Este ejemplo es solo una muestra de lo importante que son estos sistemas para las empresas que basan su negocio en la gran red de redes.

Los sistemas de recomendación llevan a cabo dos tareas principalmente: la predicción de puntuaciones y la recomendación de productos. La primera implica que el algoritmo sea capaz de predecir la puntuación (valoración dentro de un rango de valores) que un usuario daría a un producto desconocido para él. En cambio, la recomendación implica proponer un producto o productos nuevos al usuario que puedan resultar de su interés. Es fácil imaginar que lo uno lleva a lo otro: si podemos

calcular la puntuación de los productos nuevos para el usuario, entonces es posible proponer aquellos con mayor puntuación estimada.

Las estrategias seguidas para dar respuesta a estas tareas también pueden dividirse en dos: el filtrado basado en contenido y el filtrado colaborativo. Las basadas en contenido tienen en consideración exclusivamente la información propia de los productos, como su descripción, características. Por ejemplo, en el caso de una película podríamos tener en cuenta la sinopsis, el director, los actores principales, el género (acción, thriller, comedia) entre otras características. Con las películas valoradas positivamente por un usuario podemos crear un perfil en base al cual pueden seleccionarse nuevos productos.

A diferencia de esta estrategia, las estrategias de filtrado colaborativo aprovechan la información que de la comunidad de usuarios posee el sistema. Se trata de este modo de identificar usuarios con gustos similares, calculando la puntuación de nuevos productos en base a las puntuaciones de aquellos usuarios con un perfil de compra similar al nuestro, explotando así esta información colectiva.

En los métodos basados en contenido, el procesamiento del lenguaje natural resulta fundamental, pues no siempre disponemos de una descripción estructurada del producto. Esto ha llevado a proponer, entre otras, soluciones basadas en sistemas como los motores de recuperación de información, donde indexamos todas las descripciones de los productos y generamos un perfil del usuario como un agregado de todos los productos previamente asociados al mismo. Este perfil es utilizado como consulta al sistema, obteniendo así una lista ordenada de propuestas. Es una de las soluciones adoptadas en productos como las alertas de artículos científicos en Google Académico, donde a partir de los artículos de los que somos autor se construye una vector de palabras que sirve de consulta a un sistema de recuperación de información sobre los nuevos artículos indexados por Google.

Los sistemas colaborativos, en cambio, trabajan con grandes matrices usuario-producto, donde el valor para un usuario y un producto dado puede ser bien vacío (el usuario no guarda relación con el producto), bien un valor determinado como 1 o 0 para marcar si el producto ha sido adquirido por el usuario o si el usuario indicó un gusto por el producto, o incluso una puntuación sobre el mismo dentro de un rango determinado. A partir de estas matrices podemos representar a los usuarios como vectores de productos o a la inversa: representar a los productos como vectores de usuarios. En un caso u otro estaremos creando representaciones que pueden ser comparadas mediante métricas de distancia entre vectores (como la distancia euclídea, entre otras). De esta forma, dado un usuario podemos obtener la puntuación de un producto no evaluado como la media de las puntuaciones de los usuarios más próximos.

Por último, no debemos olvidar que existen sistemas híbridos que combinan ambas soluciones. Así, el valor de un elemento en la matriz usuario-producto propia del filtrado colaborativo, podría calcularse como la distancia del usuario al producto mediante una orientación basada en contenido. Algunas investigaciones han encontrado útil, por ejemplo, discriminar a los usuarios en base a sus comentarios dentro de dos categorías: optimistas y pesimistas, gracias al análisis de sentimientos sobre los textos de los comentarios, logrando una mejora cuando se aplica esta característica adicional a los sistemas de filtrado colaborativo.

Queda margen de mejora en los sistemas de recomendación y está claro que son y seguirán siendo de interés comercial. Las tendencias futuras en estos sistemas incluyen recomendaciones basadas en contexto, gracias a las tecnologías móviles, que tengan en cuenta parámetros como la hora o el lugar en el que se encuentra el usuario. También se explora el análisis de sentimientos ya comentado, así como la estimación de la reputación entre los usuarios y cuestiones adicionales que dan margen para mucha investigación por explorar.

Te animamos a completar tus conocimientos sobre los sistemas de recomendación a través de las actividades propuestas.

Referencias

Estas son las fuentes en las que me he basado para confeccionar esta lección:

- Recommender Systems Handbook, Ricci, F.; Rokach, L.; Shapira, B.; Kantor, P.B. (Eds.) 2011, ISBN 978-0-387-85819-7