

Propuesta de sistema de recomendación híbrido para sistema web de gestión y publicación de medias.

Ing. Sisley Sosa Vazquez.

Desarrollador en la Facultad 06, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

ssosa@uci.cu

Ing. Yoendry Fuentes Hernández.

Desarrollador en la Facultad 06, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

yfuentesh@uci.cu

RESUMEN.

El presente trabajo es el resumen de una investigación realizada con el objetivo de mejorar las prestaciones de un sistema web de gestión y publicación de medias elaborado en la Universidad de Ciencias Informáticas, mediante la integración de un sistema de recomendación híbrido, compuesto por el basado en contenido y el filtrado colaborativo.

El estudio fue realizado sobre un número considerable de bibliografías que abordan el tema, muestran las ventajas y problemas presentes en cada tipo, incluidas recomendaciones para corregirlos, y plantean algoritmos eficientes para la elaboración de sistemas de recomendación.

Con esta investigación se pretende formar las bases de conocimiento necesarias para comenzar la implementación de un sistema de recomendación que funcione eficientemente y arroje resultados acertados que aumenten la aceptación del sitio de publicación de medias por parte de los usuarios.

Palabras claves: basado en contenido; filtrado colaborativo; publicación de medias; sistema de recomendación híbrido.

INTRODUCCIÓN.

Para los sitios web que se encargan de publicar periódicamente audiovisuales, incrementando su mediateca con cada nueva adquisición, resultan muy necesarios mecanismos capaces de orientar a los usuarios en la selección de los mismos. Además, estos deben ser eficientes y personalizados, pues natural-

mente no todas las personas desean ver el mismo tipo de material e incluso, dentro de una misma rama de audiovisuales, los gustos son diversos.

Los sistemas de recomendación son herramientas que generan recomendaciones sobre un determinado objeto de estudio, a partir de las preferencias y opiniones dadas por los usuarios (1). Se basan en algoritmos estadísticos que intuyen los gustos de una persona tras estudiar los del resto de ciudadanos que desean un mismo producto o tras analizar otros objetos que, ya se conoce, son de preferencia para la persona (1) (2).

En todo sistema informático donde se le ofrezca al usuario un número considerable de información y/o elementos, es aconsejable que exista uno de estos sistemas, no sólo en aquellos relacionados con los audiovisuales, pero este documento es el resultado de una tarea investigativa con el objetivo de mejorar las prestaciones de un sistema web de gestión y publicación de medias, desarrollado por el equipo de trabajo del proyecto Catalogación y publicación de medias, de la Universidad de las Ciencias Informáticas (UCI), del cual los autores son miembros, y es por ello que muchos de los términos utilizados están centrados en el tema.

El principal problema con el que puede encontrarse un usuario, al enfrentarse a un sinnúmero de medias disponibles en el sistema, si no tiene una idea clara de qué desea, está dado por su incapacidad de elección, pues cuan mayor sea la cantidad de elementos más difícil se torna llevar a cabo una decisión. Entonces puede convertirse en un proceso extenuante, arrojando un resultado que seguramente es incapaz de cumplir las expectativas del sujeto, quien puede abandonar la búsqueda debido a su frustración.

Para dar solución a este problema se decidió realizar el estudio de diferentes mecanismos de recomendación que diera paso al desarrollo de una herramienta aplicable al sistema web de publicación de medias. A continuación se muestran de manera resumida los resultados del estudio.

DESARROLLO.

Flujo del proceso en sistema de recomendaciones.

Los procesos ocurridos en un sistema de recomendación pueden verse como un ciclo reiterativo, donde la información circula y se incrementa con cada iteración, pues los datos entrantes se analizan generando los resultados que a su vez servirán de realimentación para el sistema.

Como primer paso se tiene la obtención de información proveniente del usuario objetivo (entiéndase por aquel al cual se le desea recomendar), también información sobre los ítems (puede encontrarse como media o audiovisuales de ahora en adelante) y/o información del resto de usuarios del sistema, que actúan como colaboradores (1). La información de los audiovisuales puede obtenerse mediante la catalogación de los mismos, teniendo en cuenta que, entre más detallada sea la descripción, mejores resultados podrán obtenerse al re-

comendar. Por otra parte, para el caso de los usuarios existen dos métodos, en dependencia de la forma que se adopte para la creación de sus perfiles en el sistema: una es la forma implícita y la otra explícita. En el primer caso el usuario no tiene conocimiento de que sus acciones están siendo tomadas en consideración, ya sea mediante el análisis del número de visitas que recibe una publicación, el registro de los audiovisuales que el usuario ha seleccionado o visto en su ordenador. De la otra forma, el sistema brinda al usuario la oportunidad de clasificar, dentro de un rango predefinido, las medias; ejemplo de ello es solicitar al usuario que puntúe o que seleccione de entre más de un elemento cuál o cuáles son de su preferencia (1) (3).

Luego queda procesar toda la información entrada y establecer patrones que determinen, según los desplazamientos de preferencias de los usuarios colaboradores y del usuario objetivo en sí mismo, cuál audiovisual publicado en el sitio, que no haya visto, puede ser de su interés. El resultado de este análisis puede estar dado por una predicción o como una lista con los primeros elementos de todos aquellos que puede preferir el usuario (3). Finalmente, la aceptación o no del material audiovisual por parte del usuario se integrará a la información de su perfil y será utilizada para la generación de una nueva recomendación.

Clasificación de los sistemas de recomendación.

Los sistemas de recomendación pueden ser clasificados distintamente según el tipo de información que estos utilizan para elaborar la recomendación en sí (4). Entre los existentes están: el sistema de recomendación demográfico, que utiliza información demográfica obtenida en los perfiles; basado en la utilidad, que realiza cálculos de utilidad para cada objeto en particular con respecto al perfil determinado de cada usuario; basado en el conocimiento, que se apoya en inferencias sobre las necesidades de los usuarios y sus preferencias; basado en contenido, el cual determina similitud entre los ítems; y el filtrado colaborativo, que se basa en la similitud entre usuarios (4) (5) (6), siendo este último el de uso más frecuente (1) (6).

Debido a las características de la aplicación sobre el que se desea emplear el sistema de recomendación (recuérdese que está destinado a la publicación de medias), en el cual los archivos son previamente descritos con la utilización de un subsistema web de catalogación de audiovisuales, desarrollado por el mismo equipo de trabajo, y que, por su naturaleza, es de un acceso considerable de usuarios ávidos de esta forma de entretenimiento, el estudio se enfocó en los sistemas de recomendación basados en contenido y en filtrado colaborativo.

A grandes rasgos, el sistema de recomendación basado en contenido obtiene su resultado tras un análisis de comparación entre los audiovisuales que, por la información obtenida en el perfil se sabe, son de su preferencia y aquellos que no ha visto todavía. Finalmente se obtiene un resultado con las medias más parecidas a las de su gusto (4) (6). Por otra parte, los sistemas de filtrado colaborativo usan estadísticas para buscar el conjunto de los usuarios cuyas preferencias son más similares a las de un usuario determinado (vecinos cercanos)

y a continuación se combinan las preferencias de esos vecinos para producir una predicción o recomendación para el usuario (1) (3).

Sin embargo, existen algunos problemas en estos métodos, ejemplo: en el caso del basado en contenido puede incurrirse en la sobre-especialización, ya que todas las recomendaciones obtenidas serán de elementos similares aunque a veces este problema pueda corregirse agregando búsquedas aleatorias. En el filtrado colaborativo se presentan una serie de inconvenientes como el hecho de no poder hacer recomendaciones eficientes si el usuario objetivo es nuevo en el sistema y no tiene un perfil con información suficiente (7) o cuando el número de usuarios es bajo (1).

Entonces es válido mencionar otro tipo de sistema de recomendación que surge con la combinación de varios de ellos juntos: los sistemas híbridos (4) (6). En el caso de esta investigación la hibridación se plantea mediante la conjunción de las salidas respectivas del sistema basado en contenido y el de filtrado colaborativo. De esta manera pueden corregirse los errores mencionados, arrojando resultados más acertados y variados.

Solución propuesta.

La aplicación web a la cual se le quiere aplicar el sistema de recomendación, permite la creación de perfiles de usuario de manera explícita. Los usuarios tienen la posibilidad de dar puntos a los materiales que visualizan en un rango de 1 a 5 puntos, y además se le da la opción a los usuarios de elegir de cuáles tipos de audiovisuales desean recibir recomendaciones por parte del sistema.

El sistema (o subsistema) en desarrollo permite, a la administración del sistema de gestión y publicación de medias, o cualquier otro donde se aplique, seleccionar cuál de los métodos desea emplear: ya sea basado en contenido, filtrado colaborativo o un híbrido de ambos. Esto se debe a los disímiles entornos en los que el sistema puede ser desplegado.

En lugares donde los usuarios no permanecen el tiempo suficiente (el caso de los hoteles, por ejemplo) como para formar perfiles que brinden la información necesaria para formular recomendaciones acertadas, contando con la correcta descripción de los materiales, el sistema más conveniente puede ser el basado en contenido. Sin embargo, en entornos donde los audiovisuales no cuentan con estándares descriptivos o no tienen información relativa suficiente, puede ser más eficiente el filtrado colaborativo. En cualquier caso podría emplearse el sistema híbrido, pues este simplemente intentará mejorar el resultado obtenido con el primer y segundo tipo, aunque se debe tener en cuenta que esto significa dos sistemas de recomendación en ejecución, lo que pudiera repercutir en el rendimiento del sistema en general.

Para el desarrollo del sistema de filtrado colaborativo se ha propuesto utilizar el basado en memoria, en el cual se determinan cuáles son los usuarios más similares al activo y se genera una predicción para el mismo. Representando a los usuarios como vectores de las puntuaciones dadas a las medias, la similitud

entre ellos puede determinarse por el coseno que forman sus ángulos, y la predicción como una suma media ajustada (6).

En el caso del sistema de recomendación basado en contenido se determinó trabajar con la propuesta de (8), empleando redes bayesianas para cada tipo de material (video clip, película, serie, noticiero, etc.). Estas redes estarán compuestas por cuatro niveles de nodos: *a)* los términos, con el contenido del audiovisual, *b)* las categorías, que representan la estructura del contenido, *c)* los ítems, que son los audiovisuales atendiendo al tipo para el cual se construye la red, y *d)* los votos que han realizado los usuarios a las medias. De esta manera, la predicción de votación para un usuario puede determinarse por un proceso de cálculo de probabilidades que comienza en los nodos términos y termina en el nodo votación, dependiendo cada nivel de los valores calculados en el nivel anterior.

Tanto para el filtrado colaborativo como para el sistema basado en contenido, se ha determinado que las predicciones se realicen independientes para cada tipo de audiovisual, publicado en el sitio, que sea de interés para el usuario. Así se asegura que en la salida existan listas de recomendaciones para todos los tipos de materiales que el usuario ve.

CONCLUSIONES.

Después de desarrollada la investigación descrita en este artículo, se ha llegado a las siguientes conclusiones:

- La utilización de un sistema de recomendación, en el sistema web de gestión y publicación de medias de la UCI, permitirá mejorar las prestaciones que brinda el mismo, ahorrándole a los usuarios el extenuante proceso de búsqueda.
- El desarrollo de un sistema de recomendación híbrido permitirá minimizar los errores a los que se exponen el método basado en contenido y el filtrado colaborativo, y el poder elegir entre cuál de ellos ejecutar, permitirá mayor eficiencia en el sistema general para aquellos entornos de despliegue donde no sean requeridos los dos.
- El sistema de recomendación en desarrollo está soportado por algoritmos eficientes elaborados para este tipo de tareas, por lo que el resultado será, sin dudas, una herramienta que potenciará la aceptación del sistema web de gestión y publicación de medias en la red donde se despliegue.

BIBLIOGRAFÍA.

(1) Herrera-Viedma, Enrique; Porcel, Carlos; Hidalgo, Lorenzo. "Sistemas de Recomendaciones: Herramientas para el Filtrado de Información en Internet". *Hipertext.net*. Número 2. Barcelona, Universitat Pompeu Fabra, 2004.

(2) Albín Rodríguez, Antonio Pedro. *Sistema de recomendación colaborativo basado en algoritmos de filtrado mejorado*. Jaén. Escuela Politécnica Superior de Jaén. 2009.

(3) González, Alonso. *Ayudantía Sistemas de Recomendación*. Scribd. Inc. 2012.

(4) Seguido Font, Miguel. *Sistema de recomendación para webs de información sobre la salud*. Universitat Politècnica de Catalunya. 2009.

(5) Burke, Robin. "Hybrid recommender systems. Survey and experiments". *Journal. User Modeling and User-Adapted Interaction*. Número 4. 331-370. Kluwer Academic Publishers Hingham. 2002.

(6) Castellano Torres, Emilio. *Evaluación del uso de algoritmos colaborativos para orientar académicamente al alumnado en Bachillerato*. Jaén. Universidad de Jaén. 2007.

(7) Fernández Ramírez, María de Lourdes. *Ágora: Creación de Grupos Virtuales en Bibliotecas Digitales*. Cholula. UDLA. 1998.

(8) de Campos, Luis M.; Fernández-Luna, Juan M.; Huete, Juan F.; Rueda-Morales, Miguel A. "Uso de conocimiento estructurado en un sistema de recomendación basado en contenido." *Boletín de la Sociedad Cubana de Matemática y Computación*. Número Especial. 1-6. La Habana. Sociedad Cubana de Matemática y Computación. 2007.