

Sistema de recomendación por filtrado colaborativo para el sistema de publicación de contenido multimedia - VideoWeb 1.0

[Recommender system using collaborative filtering for the publication system of multimedia content - VideoWeb 1.0]

Ing. Yarisel Rojas Castellanos

Departamento GEYSED,
Universidad de las Ciencias Informáticas UCI,
La Habana, Cuba

Copyright © 2014 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: Recommender systems are software solutions that provide a list of suggestions which contains elements that can be of preference for each user. The use of these systems in multimedia content publish platforms facilitates the search of audiovisual content. The objective of this research is to analyze the techniques of collaborative filtering recommendation based on using the benefits of a large user community. In this research is carried out the description of the process flow for generating recommendations. Collaborative filtering algorithms based on users and items were analyzed and the evaluated algorithms with best results in the data set of the platform were selected. The main problems of the selected collaborative filtering technique, such as the problem of the new items, were also analyzed and solutions were proposed. The developed system was encapsulated in a module for VideoWeb 1.0 platform that uses the Drupal CMS in version 6. The results were evaluated using the mean absolute error method and presented using a range of 50 to 200 neighbors. The integration of the recommendations module to the platform provides an increase in the personalization of the multimedia content posted in order to satisfy each user preferences. This module offers an increase in the reliability of the users and minimizes the research time of the multimedia content.

KEYWORDS: drupal, personalization, neighborhood, preferences, recommender system, VideoWeb.

1 INTRODUCCIÓN

Estudios realizados en el año 2012 por la Internet Word Stats (Estadísticas mundiales de internet) arrojaron porcentajes altos del uso de la red global en las diferentes regiones del mundo [1], donde se evidenció que aproximadamente la mitad de la población conectada a la red la utiliza con fines de entretenimiento y un apreciable porcentaje para consumo de información audiovisual [2].

Diariamente surgen variedades de empresas virtuales cuyo objetivo es satisfacer las demandas de los usuarios en cuanto al comercio electrónico y el ocio, además de sacar ganancias de las mismas, entre las que se encuentran los sitios de distribución de contenido multimedia [3]. Esta tendencia permite disponer de portales web en internet con alta variedad de materiales multimedia de distintas clasificaciones, lo que provoca que la red se convierta en una nueva ventana de consumo audiovisual, en la que cada usuario elige entre las extensas ofertas de contenidos disponibles en múltiples formatos [4].

En este ámbito, los individuos se enfrentan a escenarios caracterizados por la existencia de una voluminosa cantidad de contenidos, por lo que se auxilian de las recomendaciones de personas con mayor experiencia para tratar de solucionar los

problemas derivados de su limitado conocimiento, respecto a categorías o alternativas en el consumo de materiales [5]. Como forma de automatizar este proceso, mejorar el tratamiento de los clientes a partir de propuestas personalizadas de los contenidos ofertados y satisfacer sus necesidades, surgen los sistemas de recomendación. A lo largo de los años la definición de estos sistemas ha ido evolucionando, estando estrechamente asociada a los avances de las nuevas técnicas o ideas que van surgiendo en su estudio. Algunas de sus definiciones más completas fueron formuladas en el 2005, según los autores Sung-Hwan Min, Ingo Han se conceptualiza como: *“el filtrado de información que aplica técnicas de análisis de datos para el problema de ayudar a los clientes a encontrar los productos que les gustaría adquirir, realizando una predicción de calificaciones semejantes o un listado de productos a recomendar para un determinado cliente.”* [6]

Una de las soluciones del proyecto de Catalogación y Publicación de Medias es la plataforma VideoWeb, la cual se encuentra en despliegue en su versión 1.0. Esta plataforma tiene como principal objetivo la distribución de contenido multimedia a la comunidad de usuarios de la Universidad de la Ciencias Informáticas, mediante su publicación en un portal web. La publicación de materiales responde a distintas áreas de la universidad, por ende atiende a varios intereses: educación de pregrado, educación de postgrado, auto-preparación, conocimiento general integral, entretenimiento, educación política, actualización socio-económica, información, entre otros. El sitio mantiene una distribución de los contenidos por área temática, pero aun así la cantidad de información colapsa a los usuarios en los procesos de búsquedas de un audiovisual de su interés.

El buscador de la plataforma está regido por la documentación que se tenga de cada material y los intereses de los usuarios no están enmarcados en una única área temática, sino que pueden existir audiovisuales que cumplan con sus preferencias en varios ámbitos. La escasa personalización de la información propicia un aumento del tiempo empleado en las búsquedas de un material, así como una considerable disminución de la fidelidad hacia la plataforma.

Para conformar una solución ajustada a las necesidades de la comunidad universitaria que hace uso de la plataforma y brindar una oferta personalizada para cada usuario, la presente investigación se centra en el análisis de los sistemas de recomendación. Se presenta una descripción de estos sistemas y sus clasificaciones, manteniendo como objetivo principal el análisis de técnicas que utilicen las ventajas de una cuantiosa comunidad de usuarios, específicamente la técnica de filtrado colaborativo.

El presente artículo estará organizado según la siguiente distribución: Sección II análisis y clasificación de los sistemas de recomendación, sección III presentación de trabajos relacionados, sección IV técnica de filtrado colaborativo, tendencias y problemas que presenta, sección V presentación de la metodología seguida por la investigación, sección VI Conclusiones y trabajos futuros.

2 SISTEMAS DE RECOMENDACIÓN

Los sistemas de recomendación son, a consideración de la autora, herramientas de software encargadas de tres procesos fundamentales: filtrado de información relevante, obtención de los elementos que serán recomendados y presentación de las sugerencias.

Filtrado de información relevante: se extrae información referente al usuario a partir de algunas de sus acciones sobre los contenidos o utilizando métodos directos de recopilación de preferencias (cuestionarios de preguntas, formularios). La información extraída es filtrada para eliminar los datos que no tienen suficiente peso para ser considerados una preferencia válida. De este proceso se obtiene un listado de preferencias que caracterizan los gustos del usuario.

Obtención de los elementos que serán recomendados: estos elementos se obtienen a partir de una intercepción de las características de todos los elementos existentes con las preferencias extraídas del primer proceso, para aumentar la fiabilidad se hacen uso de métodos probabilísticos que permitan cuantificar la medida en la que cada contenido multimedia cumple con el listado de preferencias, obteniendo una probabilidad de satisfacción de cada contenido para el usuario.

Presentación de las sugerencias: la presentación de las sugerencias depende de las características de cada plataforma que hace uso del sistema de recomendación. En muchos casos el sistema de recomendación se encarga de brindar un listado de sugerencias de elementos y la plataforma que hace uso del sistema se encarga de presentarlos, sin embargo en otros, un usuario administrador puede establecer una configuración para mostrar los resultados de acuerdo a sus necesidades.

La metodología a seguir para la ejecución de cada proceso varía según la técnica de recomendación utilizada y está estrechamente relacionada al tipo de elemento que se recomienda. La presente investigación se enfoca en recomendaciones

de contenidos multimedia, estos elementos presentan varias características que los diferencian de otros elementos, como la cantidad de descriptores que pueden ser insertados para su caracterización.

Existen varias técnicas utilizadas en los sistemas de recomendación, entre las más abordadas por la comunidad científica se encuentran:

- **Sistemas de recomendación basados en contenido:** recomienda materiales audiovisuales similares a aquellos que les han gustado al usuario, en función de características de los mismos [7].
- **Sistema de recomendación por filtrado colaborativo:** recomienda contenidos multimedia que hayan tenido votaciones o valoraciones satisfactorias y que no hayan sido vistos por el usuario, pero sí por usuarios que comparten preferencias similares a este [8].
- **Sistemas de recomendación basados en conocimiento:** realiza sugerencias de audiovisuales fundamentadas en inferencias sobre las necesidades de los usuarios y sus preferencias [9]. Estos sistemas están regidos por conocimientos del entorno en el que se generan las recomendaciones y es provisto por especialistas en la materia.
- **Sistemas de recomendación basados en la utilidad:** utiliza una fórmula de utilidad que interviene directamente en el proceso de la recomendación y que contempla variables como disponibilidad, calidad de los contenidos y formatos.
- **Sistemas de recomendación demográficos:** realiza las recomendaciones a partir de información geográfica, lo que permite crear grupos geo-localizados de usuarios a partir de su consumo audiovisual.

Cada una de las técnicas anteriores tiene sus propias fortalezas, pero presentan problemas como la alta dependencia de conocimiento del entorno o el requerimiento de gran cantidad de datos que caractericen a los materiales que recomiendan, por citar algunos; para minimizar estos problemas y aumentar la fiabilidad de las recomendaciones surgen los sistemas híbridos. Estos sistemas son formados por dos o más técnicas de recomendación a partir de varios criterios de hibridación. Los sistemas híbridos tienen altos indicadores de precisión en las recomendaciones pero son complejos en su desarrollo.

3 TRABAJOS RELACIONADOS

A partir del auge de los sitios de distribución de contenido multimedia un considerable número de sistemas de recomendación han sido introducidos para mejorar el tratamiento de los usuarios. Entre ellos se encuentra el sistema de recomendación de FilmAffinity [10] que recomienda contenidos multimedia asociados al área del cine. Este sistema cuenta con una base de datos con fichas técnicas detalladas de cuantiosa cantidad de películas, documentales, cortometrajes y series de televisión. Su funcionamiento se basa en el cálculo de la media de puntuaciones realizadas por los usuarios tanto a cada contenido como a sus críticas.

El sistema de recomendación películas de Movielens [11] está basado en técnicas de filtrado colaborativo, este sistema ha sido desarrollado por el GroupLens Research de la Universidad de Minnesota. Cada usuario del sistema tiene asociado una vecindad de usuarios afines calculada a partir de las opiniones emitidas sobre las películas. La calificación de estos vecinos se utiliza para generar las recomendaciones.

Last.fm [12] es un sistema de recomendación de música. Obtiene las preferencias de los usuarios con gráficos generados a partir de las canciones más reproducidas por cada usuario y los autores asociados a estas. A partir de estos gráficos se calculan las recomendaciones finales haciendo uso de técnicas de filtrado colaborativo, esto permite a los usuarios explorar listas de reproducción de usuarios con preferencias similares.

Cada uno de los sistemas analizados presenta características específicas según los datos que utilizan para la generación de las recomendaciones, así como las acciones que deben realizar los usuarios sobre los contenidos multimedia para obtener sus preferencias. Analizándolos por separado se concluyó que ninguno cumple con las necesidades de la plataforma VideoWeb 1.0, debido a la gestión de tipologías que realiza la plataforma para cada contenido.

4 TÉCNICA DE FILTRADO COLABORATIVO

El filtrado colaborativo es una de las técnicas más utilizadas en los sistemas de recomendación. Está enfocado en la predicción y recomendación de elementos a partir de la recolección de preferencias de muchos usuarios. Se basa en la premisa de que si dos o más usuarios tienen similares características (comportamiento en cuanto a sus votaciones) deben

tener similares preferencias. Esta medida de similitud es cuantificada para asociar a cada usuario un grupo de colaboradores¹.

Existen dos tendencias para los algoritmos de filtrado colaborativo: los basados en memoria y los basados en modelos. Los **basados en memoria** realizan las predicciones a partir de una base de datos en la que se almacenan los usuarios, los contenidos multimedia y las votaciones con sus respectivas relaciones. Con estos datos se obtiene el historial de valoraciones de los colaboradores del usuario actual y se infiere una lista de audiovisuales para ser recomendados. Los **basados en modelos** crean un modelo de los usuarios y sus valoraciones, a partir de este se calcula el valor esperado para cada multimedia en función de las valoraciones realizadas por los colaboradores. Se basa en algoritmos de aprendizaje [13].

Los algoritmos de recomendación colaborativa basados en memoria establecen las relaciones de similitud entre usuarios o entre elementos. Estos algoritmos se basan en fórmulas matemáticas para cuantificar la semejanza entre dos de estas categorías. La técnica de filtrado colaborativo presentan varios problemas, entre estos se encuentra la escasez de datos debido a la necesidad de una cuantiosa cantidad de usuarios realizando valoraciones de contenidos similares, estos datos deben proveer una matriz de puntuaciones que permita realizar los cálculos de vecindad, predicciones y recomendaciones con un alto valor de fiabilidad. Otro problema es la escalabilidad del sistema: dada porque el costo computacional de los algoritmos utilizados para calcular los vecinos más cercanos crece a medida que aumenta el corpus de datos.

Basado en el funcionamiento de esta técnica un contenido multimedia es recomendado al usuario activo si alguno de sus colaboradores ha realizado valoraciones satisfactorias sobre este, cuando se publica un nuevo contenido en la plataforma este no es recomendado hasta que no haya recibido alguna valoración, presentando así el problema del nuevo elemento. Al registrarse un nuevo usuario en la plataforma no se tiene información suficiente sobre sus preferencias, debido a que no ha realizado valoraciones de los contenidos publicados, como consecuencia no puede ser asociado a ningún grupo y no recibe recomendaciones [14].

5 PROPUESTA DE SOLUCIÓN

Para la selección de la tendencia colaborativa, utilizada en la solución propuesta, se analizaron las características de la plataforma VideoWeb 1.0 que podrían interactuar directamente con cada uno de estos métodos. Los algoritmos de filtrado colaborativo basados en modelos tienen buenos resultados de precisión en entornos que manejan gran cantidad de datos [15], teniendo en cuenta que la cantidad de usuarios registrados en la plataforma y la cantidad de contenidos multimedia que se gestionan en la misma es muy variable, la generación de un modelo puede ser poco propicia en varios casos presentando el problema de la escasez de datos. En correspondencia a las características de los algoritmos basados en memoria es sustancial la utilización de la base de datos con que cuenta la plataforma. Para calcular la similitud se le presenta mayor interés a los basados en elementos, aunque se desarrollan ambos algoritmos para comparar los resultados obtenidos por estos en las pruebas realizadas con el corpus de datos que utiliza la plataforma.

5.1 METODOLOGÍA SEGUIDA POR LA INVESTIGACIÓN

La metodología utilizada para el desarrollo de la investigación cuenta con cuatro fases fundamentales, de las que se derivan varios procesos asociados al funcionamiento del módulo desarrollado. En la primera fase se establecen los procesos referentes a la detección de los datos necesarios para la generación de recomendaciones. En la segunda fase se especifica el funcionamiento del algoritmo para el cálculo de vecindad y la medida de similitud utilizada. En la tercera fase se presenta el flujo de las recomendaciones en la plataforma, se establecen soluciones para los problemas persistentes en el algoritmo de filtrado colaborativo desarrollado y se especifican características de la construcción de la solución. La última fase se encarga de la presentación y evaluación de los resultados obtenidos.

¹ Colaboradores: es la denominación que se les da a los usuarios en el filtrado colaborativo refiriéndose a los que comparten preferencias similares al usuario analizado, matemáticamente son los que su grado de similitud están más próximos al 0 en un rango de (0,1].

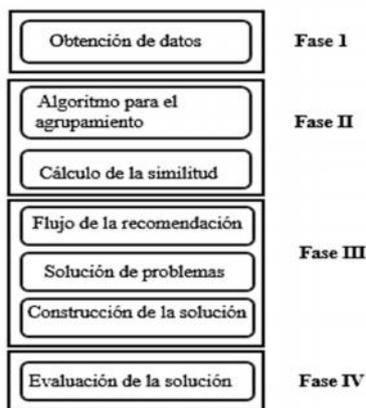


Fig. 1. Metodología utilizada en la investigación.

5.2.1. OBTENCIÓN DE DATOS

Las acciones realizadas por los usuarios sobre los contenidos multimedia publicados en la plataforma son almacenadas en base de datos por los módulos que la componen. Para la generación de las recomendaciones se realiza una extracción de las tablas y los datos necesarios para la ejecución de los procesos. Estos datos fueron el identificador, nombre, título, categoría y descriptores de las publicaciones de archivos multimedia, los usuarios registrados en la plataforma con los datos asociados a estos presentes en la tabla usuario y las votaciones almacenadas por el módulo *voting_api*, de estas votaciones se obtuvo el valor del voto, el usuario que la realizó, y el audiovisual al que está asociada.

5.2.2. ALGORITMO PARA EL CÁLCULO DE VECINDAD

Los datos recolectados en el proceso anterior son analizados para su agrupación por similitud, en la investigación se desarrollaron el cálculo de la similitud basados en usuarios y basados en contenidos multimedia. La explicación de los procesos se presenta a partir del cálculo de la similitud basados en contenidos multimedia debido a que ambos flujos de eventos son muy similares. El cálculo de vecinos más cercanos se realiza a partir del siguiente flujo:

- I. Se eligen aleatoriamente K contenidos multimedia que formarán los centroides² de los K grupos iniciales: la cantidad de grupos (K) será un dato especificado por el administrador en la configuración del módulo, debido a que esta cantidad debe variar en correspondencia con la cantidad de usuarios registrados y audiovisuales publicados.
- II. Los audiovisuales asignados como centroides son eliminados del conjunto total.
- III. Se calcula la similitud entre cada contenido multimedia y los centroides seleccionados para cada grupo, cada contenido será asignado al grupo del centroide con el que comparte menor similitud.
- IV. Cuando todos los audiovisuales han sido asignados a un grupo se vuelven a calcular los centroides por cada grupo, este paso se realiza a partir del cálculo de la similitud de cada contenido multimedia con los restantes del grupo, tomando como nuevo centroide al de menores similitudes totales.
- V. Cuando se han seleccionado los nuevos centroides se vuelven a ejecutar los pasos del II al IV, estos pasos se ejecutan hasta que los centroides anteriores son iguales a los obtenidos después del paso IV.
- VI. Se almacenan los grupos en la base de datos y los contenidos multimedia asociados a cada uno.

² Centroides: se refiere a los elementos de mayor generalización del grupo, inicialmente se toman de forma aleatoria, pero luego de varias iteraciones del algoritmo son actualizados por los más relevantes de cada grupo.

5.2.3. CÁLCULO DE LA SIMILITUD

En los pasos III y IV del algoritmo anterior se ejecuta el cálculo de la similitud entre contenidos multimedia a partir de la distancia euclidiana. Esta métrica de similitud se puede expresar como la distancia entre dos contenidos multimedia, partiendo de los usuarios como las dimensiones entre estos y sus valoraciones, los pesos a lo largo de esas dimensiones. El cálculo se realiza con los usuarios (dimensiones) que han realizado votaciones por ambos contenidos. En un lenguaje matemático se puede expresar como la raíz cuadrada de la sumatoria (suma por cada dimensión) del cuadrado de la diferencia de votaciones para cada par de contenidos.

$$\sqrt{\sum_{n=0}^n (N_i - M_j)^2}$$

Ecuación 1: Distancia euclidiana entre dos contenidos multimedia.

Donde:

m: cada uno de los usuarios (dimensión).

n: cantidad de usuarios que han votado por ambos contenidos multimedia.

i, j: contenidos multimedia para comparar su similitud.

M_i: votaciones realizadas por el usuario m al contenido multimedia i. (Matriz de valoraciones)

M_j: votaciones realizadas por el usuario m al contenido multimedia j. (Matriz de valoraciones)

5.2.4. FLUJO DE LA RECOMENDACIÓN

Para generar una recomendación en la plataforma se parte de la autenticación del usuario, seguidamente se busca el grupo al que se asocian los contenidos multimedia vistos por el este. Seleccionado el grupo se calcula la valoración de los contenidos multimedia pertenecientes al grupo pero que no han recibido valoraciones por el usuario. Se recomiendan los diez audiovisuales de mejor valoración. El flujo de eventos completos se ejecuta según el diagrama mostrado en la Fig. 2.

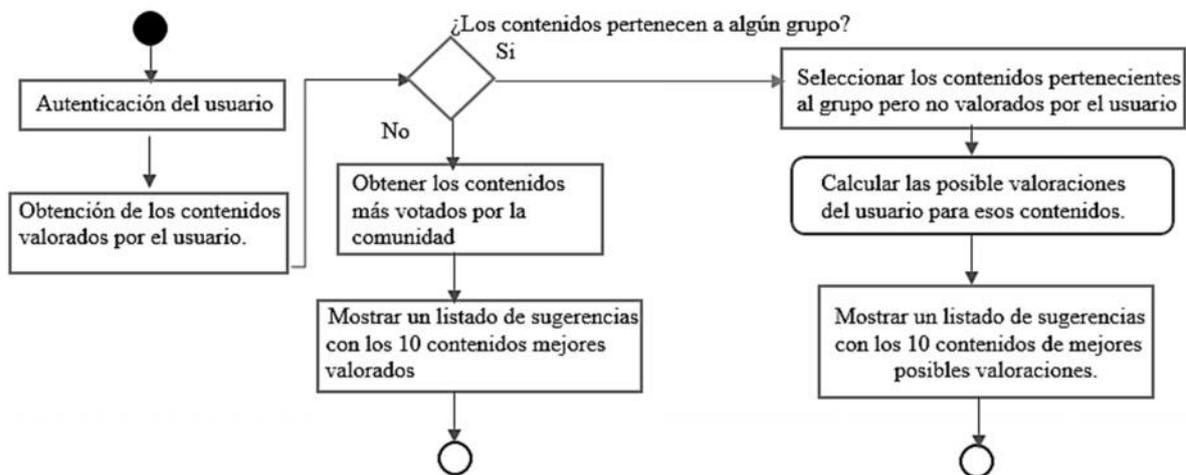


Fig. 2. Flujo de eventos para generar recomendaciones.

5.2.5. SOLUCIÓN DE PROBLEMAS DE LA TÉCNICA IMPLEMENTADA

Según las características del entorno en el que se utiliza la plataforma VideoWeb 1.0 el sistema de recomendación desarrollado no presenta el problema de la escasez de datos, debido a la cuantiosa comunidad de usuarios y contenidos multimedia publicados. El problema de la escalabilidad fue minimizado a partir de la utilización de un algoritmo basado en memoria que calcula la similitud a partir de los contenidos multimedia publicados, debido a que su cantidad es en gran medida menor que la cantidad de usuarios registrados.

Para minimizar el problema del contenido multimedia nuevo la plataforma muestra una recomendación con las últimas publicaciones de contenidos multimedia, aumentando las posibilidades de que se realicen votaciones a una nueva publicación de audiovisual. Como estrategia para la generación de recomendaciones a usuarios nuevos, teniendo en cuenta que no se tiene información de sus preferencias, se toman como colaboradores todos los usuarios registrados y se presentan como sugerencias los contenidos más votados por la comunidad.

5.2.6. CONSTRUCCIÓN DE LA SOLUCIÓN

El sistema de recomendación desarrollado está encapsulado en un módulo para la versión 6 del sistema de gestión de contenidos Drupal, debido a que en este CMS se desarrolló la plataforma VideoWeb 1.0. Se utilizan algunas de las funcionalidades del núcleo del CMS definidas en el componente .module. La solución cuenta con una interfaz administrativa en la que un usuario, con permisos de administración, puede definir la cantidad de grupos necesarios para el cálculo de vecindad, una vez que se ha instalado el módulo en la plataforma. Los componentes y sus relaciones están regidas por la familia de estilos arquitectónicos de llamada y retorno, haciendo uso del patrón arquitectónico Modelo-Vista-Controlador.

La integración con la plataforma se realiza a partir del módulo bloques, que es el encargado de mostrar a los usuarios los bloques de contenido en las distintas interfaces, esto se realiza a partir de los selectores de datos que implementa. Para esta integración se desarrolló un selector de recomendaciones que obtiene un listado de 10 sugerencias de contenido multimedia, este selector se comunica con el módulo de filtrado colaborativo desarrollado haciendo uso de una de la funcionalidad del núcleo de Drupal definida para ello (module_invoke).

5.2.7. EVALUACIÓN DE LOS RESULTADOS

Para definir la fiabilidad de las recomendaciones fue necesaria la utilización de métricas de evaluación, que permitiesen medir la calidad de los resultados obtenidos. Encontrar un estándar para la evaluación de sistemas de recomendación es un problema aún sin resolver en la comunidad científica, por ello los métodos de evaluación desarrollados varían para cada técnica de recomendación. La métrica utilizada para evaluar los resultados de la presente investigación fue el Error absoluto Medio (MAE), esta métrica tiene buen impacto en la evaluación de métodos estadísticos.

El principal objetivo de las pruebas fue la comparación de las potencialidades del cálculo de la similitud entre usuarios y entre contenidos multimedia en base a las variables que intervienen en el proceso. Para la evaluación de los resultados se instaló el módulo en una distribución local de la plataforma VideoWeb 1.0 y se probó el algoritmo desarrollado para el conjunto de datos obtenidos del despliegue de la plataforma en la Universidad de las Ciencias Informáticas en período de prueba. El conjunto de datos tomados para las pruebas fueron 950 votaciones realizadas por 600 usuarios a 162 contenidos multimedia, estos fueron separados en un 70% de entrenamiento del sistema de recomendación y un 30% de pruebas. Los valores de vecinos seleccionados para las pruebas fueron entre 50 y 200 vecinos en un rango de 50, estos valores fueron tomados a partir de los datos del conjunto de prueba.

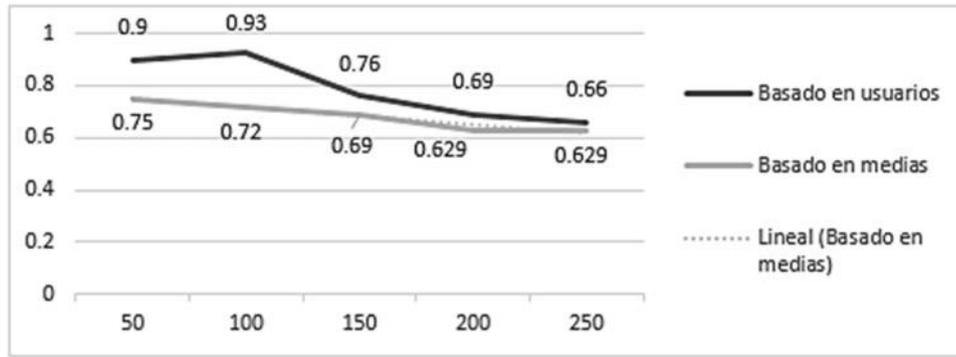


Fig. 3. Gráfica que muestra los resultados de MAE de ambos algoritmos a partir de los números de vecinos seleccionados.

En la Fig. 3 se representa el valor del MAE en el eje vertical y en el eje horizontal la cantidad de vecinos seleccionados para cada prueba, estos valores son representados para el algoritmo de filtrado colaborativo basado en usuarios y para el algoritmo de filtrado colaborativo basado en contenidos multimedia. Como se muestra en el gráfico los valores de MAE oscilan entre 0,69 y 0,93 con un número de vecinos de 50 a 150, esto indica que las recomendaciones no son muy buenas debido a que se debe aumentar el número de vecinos para agrupar los usuarios con características similares. A partir de 150 a 200 vecinos se muestra una tendencia de decrecimiento del MEA al aumentar el valor de la vecindad. Las pruebas realizadas para más de 200 vecinos se mantuvieron con una tendencia muy aproximada a los valores de MEA para 200 vecinos.

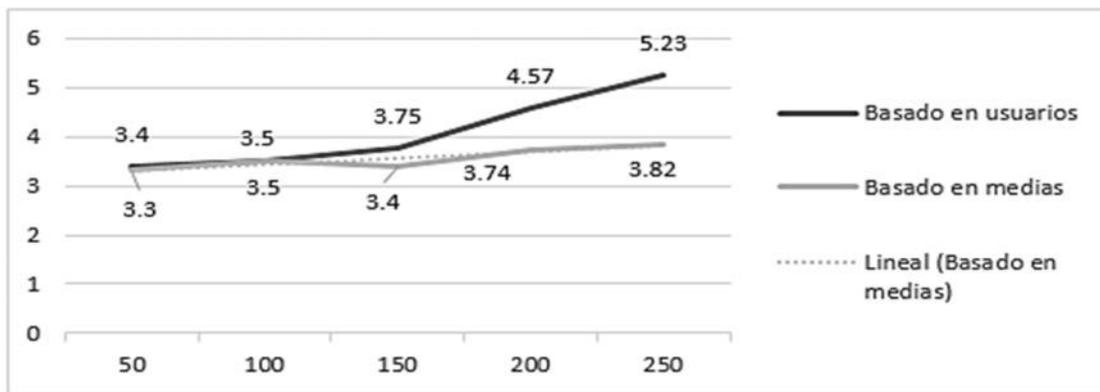


Fig. 4. Gráfica que muestra los resultados del tiempo de ejecución de ambos algoritmos a partir de los números de vecinos seleccionados.

En la Fig. 4 se presenta, en el eje vertical, el tiempo de ejecución de cada algoritmo para las diferentes vecindades representadas en el eje horizontal. Como se muestra en el gráfico el tiempo de ejecución de ambos algoritmos crece a medida que se aumenta la cantidad de vecinos, teniendo tiempos similares para una vecindad entre 50 y 150 usuarios. Se puede apreciar que a partir de 150 a 250 vecinos los tiempos de ejecución del algoritmo basado en usuarios tienen variaciones de mayor magnitud en comparación con los del algoritmo basado en contenidos multimedia.

Las pruebas realizadas indicaron un mejor comportamiento de los algoritmos desarrollados para una vecindad de 200 usuarios, teniendo en cuenta la métrica del valor absoluto medio. Los valores de MEA para el algoritmo basado en medias arrojaron mejores resultados que el algoritmo basado en usuarios. Analizando los tiempos de ejecución para la vecindad seleccionada de ambos algoritmos se evidencian mejores indicadores en el algoritmo basado en medias.

6 CONCLUSIONES Y TRABAJOS FUTUROS

La personalización de la información es una tendencia altamente utilizada por las empresas emergentes en los mercados en línea, enfocado a los sitios de distribución de contenido multimedia se ha convertido en una forma para promover el consumo audiovisual. En este sentido, la integración del módulo de recomendación a la plataforma VideoWeb 1.0 permite

presentar los contenidos publicados de acuerdo a las preferencias particulares de cada usuario, lo que le propicia un valor agregado a la misma viabilizando el proceso de búsqueda de un material y brindando un servicio ajustado a las necesidades de cada consumidor. Los resultados alcanzados por las pruebas posibilitaron la selección del cálculo de la similitud basado en elementos por presentar mejores indicadores en cuanto al MAE y el tiempo de ejecución en el entorno en el que se desarrolla la plataforma, esta selección estuvo guiada además por las tendencias del crecimiento del corpus de datos manipulado en el negocio de VideoWeb 1.0.

La técnica de filtrado colaborativo presenta buenos resultados, pero sus indicadores pueden ser mejorados a partir de su hibridación con otras técnicas. Se propone como trabajos futuros el análisis de otras técnicas de recomendación ajustadas al negocio de la plataforma VideoWeb 1.0. Además, se propone investigar sobre las técnicas de hibridación que arrojen mejores resultados frente a las pruebas realizadas al módulo desarrollado. La hibridación del filtrado colaborativo con otra técnica de recomendación es una tendencia muy utilizada en el mundo, esta estrategia se enfoca en el fortalecimiento de estos algoritmos con otros que minimicen las deficiencias que presentan las técnicas colaborativas. Para minimizar la escasez de los datos presentes en la matriz de votaciones se propone el análisis del historial de navegación del usuario teniendo en cuenta acciones sobre los contenidos multimedia como descarga o reproducción de un material, así como los comentarios realizados a una publicación; estos elementos pueden ser normalizados e incluidos en la matriz de valoraciones.

REFERENCIAS

- [1] IWS. Internet Word Stats. *Internet Usage Statistics*. [En línea] 30 de 06 de 2012. [Citado el: 26 de 10 de 2012.] <http://www.internetworldstats.com/stats.htm>.
- [2] Nielsen. Nielsen. *Reports-Media and Entertainment*. [En línea] 25 de 02 de 2014. [Citado el: 26 de 02 de 2014.] <http://www.nielsen.com/us/en/reports/2014/u-s-home-entertainment-industry-year-end-review-2013.html>.
- [3] Buendía, Lola. Nuevas formas de distribución de contenido audiovisual. *Nuevas Tecnologías by Suite101*. [En línea] 12 de 08 de 2010. [Citado el: 26 de 12 de 2013.] <http://www.nuevatecnologias/nuevas-formas-de-distribucion-de-contenido-audiovisual-a23117.html>.
- [4] Peña, Fernández Emilio. Tendencias en el consumo audiovisual. *Cambios en los medios tradicionales: participación y personalización*. [En línea] 07 de 03 de 2007. [Citado el: 26 de 10 de 2012.] <http://buscarencontrar.blogspot.com/2007/03/tendencias-en-el-consumo-audiovisual.html>.
- [5] Cruz, Constanza Claudia. Sistema de recomendación en el Comercio Electrónico y la E-educación. Bogotá (Colombia) : s.n., 2010. Vol. Vol. 8, Criterio Libre N-12. ISSN 1900-0642.
- [6] Han, S.H y Min, I. Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Applications*. 2005. Vol. vol. 28.
- [7] Campos, Luis M. y Fernández, Juan M. *Uso de conocimiento estructurado en un sistema de recomendación basado en contenido*. Andalucía : s.n., 2011. TIN2005-02516.
- [8] Salazar, J Erika. Sistema de Búsqueda Personalizada y Recomendación de Documentación Científica. Universidad de Antioquia, Medellín, Colombia : Revista Iberoamericana de Inteligencia Artificial, 2006. Vol. vol.1, No.30. ISSN: 1137-3601.
- [9] Font, Miguel. *Sistemas de recomendación para webs de información sobre la salud. Tesis de Maestría*. Madrid : Universidad Politécnica., 2009.
- [10] Fera, Liset. *Desarrollo de un sistema de recomendación colaborativo para la plataforma VideoWeb*. La Habana, Cuba : s.n., 2012.
- [11] García, Verónica y Duarte, Antony. *Análisis y Comparación de Algoritmos con Filtro Colaborativo para Motor de Recomendación utilizando Mahout. Ingeniero en Ciencias Computacionales, especialización en sistemas*. Guayaquil, Ecuador : s.n., 2011.
- [12] Betarte, Leticia, Molina, Valeria y Machado, Rodrigo. *PGMúsica Sistema de recomendación de música*. Universidad de la República : Instituto de Computación, 2006.
- [13] Ibáñez, Vázquez Felipe. *Métodos de agrupamiento en el cálculo de distancias entre usuarios en un sistema de recomendación basado en algoritmos de filtrado colaborativo*. Leganés : s.n., 2011.
- [14] Parsons, Jeffrey , Ralph, Paul y Gallagher , Katherine. *Using Viewing Time to Infer User Preference in Recommender Systems*. Canada : Faculty of Business Administration, 2004.
- [15] Cacheda, Fidel, y otros. Acta del I Congreso Español de Recuperación de Información. *Eficiencia y precisión de algoritmos de filtrado colaborativo: análisis y comparativa*. Madrid, España : s.n., 2010.