

Adaptación de un Recomendador de Filtro Colaborativo Basado en el Usuario para la Creación de un Recomendador de Materias de Pregrado Basado en el Historial Académico de los Estudiantes

Del Pino, J.; Salazar, G.; Cedeño, V. Msc.
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral
Campus “Gustavo Galindo Velasco” La prosperina Km 30,5 vía perimetral, Guayaquil, Ecuador
{ jdelpino,gsalazar,vcedeno }@fiec.espol.edu.ec

Resumen

Los sistemas de recomendación son ampliamente utilizados hoy en día gracias a su capacidad de analizar las preferencias de usuarios y sugerir ítems. No obstante, el uso de los recomendadores está limitado a un modelo basado en el usuario y no en su historial de preferencias, discriminando así el campo de aplicación, por ejemplo, a sistemas académicos donde sea primordial el estudio de las decisiones del estudiante a lo largo de su carrera. El presente trabajo presenta un esfuerzo por adaptar filtros colaborativos basados en el usuario a filtros colaborativos basados en el historial del usuario. Con un conjunto de pruebas mediremos su efectividad utilizando dos algoritmos distintos de similitud para recomendar materias a un estudiante en el sexto semestre de la carrera de Ingeniería en Electrónica y Telecomunicaciones ofertada por la FIEC - ESPOL. Los resultados muestran que es factible adaptar un recomendador a un modelo basado en el historial del usuario.

Palabras Claves: Sistema de recomendación, Recomendador de materias, Filtro colaborativo, Guía académica.

Abstract

Recommender Systems are nowadays broadly used thanks to their potential on analyzing user preferences and suggesting items to them. However, the use of Recommenders is also limited to a user based model and not to a history preferences based model. The present work represents an effort to adapt a collaborative filter based on users to a collaborative filter based on history preferences of users. A set of tests will be conducted to measure its effectiveness, using two different similarity algorithms to recommend courses to a student for the sixth semester of the Electronics and Telecommunications career offered by FIEC - ESPOL. The results show that it is feasible to adapt a Recommender to a model based on the user's history preferences.

Keywords: Recommender system, Subject recommender, Collaborative Filter, Academic guidance.

1. Introducción

Durante los últimos años, los sistemas de recomendación han sido herramientas que han contribuido a mejorar la experiencia del usuario. El conocimiento de los gustos de los usuarios ha motivado a las empresas a generar sugerencias personalizadas para agregar valor a la interacción que tienen con el contenido mostrado. La gran cantidad de información filtrada de algunos usuarios sirve para ayudar a otros, resultando en un entorno colaborativo.

Los sistemas de recomendaciones son ampliamente usados para fines comerciales: Amazon, Facebook, Twitter, Netflix, son algunos ejemplos. Sin embargo, no han sido explotados de la misma manera con fines académicos. En el campo educativo, corresponden en su mayor parte a la recomendación de objetos de aprendizaje, dentro de sistemas de administración de aprendizaje (LMS por sus siglas en inglés), también son

usados en la recomendación de recursos dentro de las librerías científicas digitales [1], [2] y [3]. Es escaso su uso para predecir y recomendar las materias que tomarán los alumnos en una institución educativa ya que en la mayoría de ellas el flujo de materias es fijo y el estudiante se ve obligado a seguirlo, con excepción de las materias opcionales. El caso más común de materias fijas y obligadas a seguir es el de las que debe tomar un estudiante de secundaria. Sin embargo, en varias universidades, aunque exista un flujo de materias con sus pre-requisitos y co-requisitos, hay más libertad para tomar la decisión de cual materia elegir y registrarse.

El presente artículo introduce una solución informática que permite adaptar un Recomendador basado en el usuario de Mahout, una librería de código abierto, para realizar recomendaciones a la situación de requerir sugerencias de registros basadas en experiencias pasadas de otros estudiantes, mientras se

analizan dos métricas de similaridad: *Tanimoto* y *LogLikelihood* [4].

El contenido de este artículo se distribuye de la siguiente manera: en la sección 2 se discute la motivación que llevó a escribir el artículo. En la sección 3 se describe el trabajo relacionado, seguida de la sección 4, dedicada a definir conceptos básicos sobre sistemas de recomendaciones. En la sección 5 se describe la metodología empleada para el desarrollo de la solución. La sección 6 incluye el análisis de los resultados. Finalmente, en la Sección 7 concluimos y detallamos el trabajo futuro.

2. Motivación

En el caso de la Escuela Superior Politécnica del Litoral (ESPOL), los estudiantes tienen la libertad de tomar materias en el orden, horario y con el profesor que el estudiante elija. Debido a ello, se observa que para estudiantes que entraron en un mismo año, tienen opciones de registros distintas en cada semestre. Esto colabora a crear matices en las rutas que toman los estudiantes a lo largo de sus carreras.

Recomendar un orden en el que un estudiante toma las materias de su carrera semestre a semestre es una utilidad muy provechosa de esta tecnología. Existen diversas formas de implementar un recomendador. Para el caso particular de recomendadores basados en el usuario, las experiencias exitosas previas de otros alumnos en el orden en que estos han aprobado las materias de su carrera, servirán de apoyo para asistir un estudiante, al momento de tomar decisiones previo a un nuevo registro de semestre en la ESPOL [4].

Uno de los retos encontrados para la implementación de un recomendador para este caso, es que no es suficiente conocer las decisiones de los estudiantes sino también el momento (semestre) en que fueron realizadas.

En la actualidad existen sistemas que recomiendan materias optativas para asistir a los asesores en la tarea de orientar a los estudiantes [5], no obstante entre las variables que se toman en consideración para realizar la recomendación, no está incluido el momento (tiempo) en el que otros estudiantes tomaron dichas materias, ya que la institución (colegio) define cuando deben ser cursadas. En el caso particular de la Escuela Superior Politécnica del Litoral (ESPOL), el momento en que fueron tomadas las materias es fundamental para la recomendación dado que los estudiantes tienen libertad de tomar diferentes rumbos en su flujo.

3. Trabajos relacionados

En nuestro trabajo exploramos la posibilidad de hacer recomendaciones de materias para estudiantes previo a su registro en un nuevo término académico. Trabajos previos usan redes neuronales y clasificadores de árbol de decisión para diseñar un sistema de

recomendaciones que basado en el historial académico de estudiantes de secundaria recomienda a estudiantes las universidades en las cuales pueden ser admitidos [6]. Otras implementaciones son las de AACORN [7] que recomienda cursos usando un sistema de recomendación basado en casos [8], un subconjunto de los basados en contenido, y RARE [9] que usa reglas de asociación e información sobre cursos que han tomado estudiantes de maestría para recomendar cursos a los nuevos alumnos. CourseAgent recomienda cursos usando un enfoque basado en la navegación que realiza en la comunidad que utiliza la aplicación en la Escuela de las Ciencias de la Información de la Universidad de Pittsburgh [10]. O'Mahony y Smyth proponen la combinación de sistemas de filtro colaborativo y basados en contenido para recomendar materias electivas a estudiantes universitarios. Un estudio similar realizan Castellano y Martínez usando filtros colaborativos para recomendar materias de especialización y electivas en el sistema de escuelas secundarias español [4].

No obstante, en ninguno de los casos mencionados anteriormente se ha implementado un recomendador basado en el usuario que realice recomendaciones de materias basado en una historia de registro, como el caso de estudio del presente trabajo.

4. Conceptos y Tecnologías Empleadas

- **Modelo de Datos.** Representación de un repositorio de información sobre usuarios y preferencias asociadas a ítems.
- **Similaridad.** Representación numérica del grado de semejanza entre dos usuarios, basada en las preferencias sobre ítems comunes.
- **Métrica de similaridad.** Medida que determina el grado de similaridad entre dos objetos.
- **Estimación de preferencia.** Cálculo que determina un valor para el grado de preferencia de un usuario por un ítem.
- **Vecindad.** Conjunto de usuarios semejantes, construido en base a las similitudes de los usuarios.

5. Materiales y Método

5.1. Adaptación del Recomendador

El método propuesto para la adaptación del recomendador basado en el usuario consiste en introducir un nuevo concepto de vecindad. Esta vecindad, que se ha definido como *Vecindad por Historial*, contiene el conjunto de estudiantes que mantienen la mayor semejanza en el orden de aprobación de materias con respecto al estudiante que recibe la recomendación, al cual lo llamaremos de ahora en adelante como *Estudiante E*. Esta vecindad es

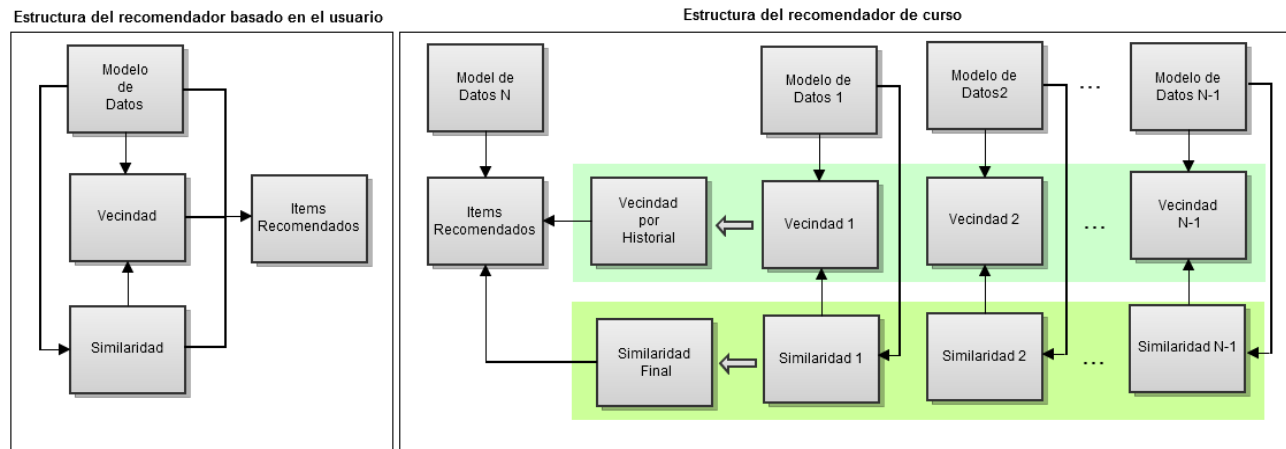


Figura 1. Elementos del recomendador basado en el usuario (izquierda). Elementos del recomendador de curso basado en historial académico (derecha)

el producto de un recorrido en el historial académico del *Estudiante E*. Para recomendar materias en un Semestre N , se calculan las vecindades individuales para cada uno de los semestres anteriores a N , es decir, desde el primer semestre, hasta el semestre $N-1$. Por cada Semestre S_i , se establece una Vecindad V_i , obtenida utilizando las funciones sin modificar de la librería Mahout. Cada Vecindad V_i contiene un conjunto de estudiantes que mantienen un grado de similaridad con el *Estudiante E* en un Semestre S_i , con respecto a las materias tomadas en ese semestre.

Con el conjunto de Vecindades V_i , se construye una *Tabla de Frecuencia (Hashmap)* cuya estructura contiene: los IDs de los estudiantes (*Key*), y un valor de repetición (*Value*). El objetivo de la *Tabla de Frecuencia* es registrar el número de semestres en los que cada uno de los estudiantes evaluados mantuvo una similaridad con el *Estudiante E*. Dicha similaridad está determinada por la métrica de similaridad utilizada en el cálculo. En otras palabras, la *Tabla de Frecuencia* puede ser vista como una lista ordenada en forma descendente basada en el puntaje de aparición de un estudiantes en las vecindades V_i .

Una vez obtenida la *Tabla de Frecuencia*, es posible obtener la *Vecindad por Historial*. Para su efecto, se utilizan diversos métodos. En nuestro trabajo, hemos considerado dos: *Threshold* y *Nearest-N* [4]. De ellos depende la cantidad de Estudiantes filtrados a la *Vecindad por Historial*.

Paralelamente a la construcción de la *Vecindad por Historial*, se calcula la *Similaridad Final* de cada uno de los estudiantes pertenecientes a la *Tabla de Frecuencia*.

La *Similaridad Final* es el promedio de todos los valores de similaridad que un estudiante registró en los Semestres S_i con respecto al *Estudiante E*.

La *Similaridad Final* es usada en el proceso de *estimación de preferencia* de una materia, de esta manera las materias que los estudiantes de la *Vecindad por Historial* aprobaron en su Semestre N , son los cursos que se entregan como recomendación al

Estudiante E.

La Figura 1 muestra las diferencias entre el método de recomendación tradicional de Mahout (izquierda) al método modificado y propuesto en el presente trabajo (derecha). Como se puede observar en la gráfica, Mahout realiza la recomendación de ítems basado en un *Modelo de Datos (Datamodel)* único, para posteriormente obtener una única Vecindad luego de establecer Similaridades. En contraste, en el trabajo propuesto, los ítems a recomendar provienen del *Modelo de Datos* que representa el Semestre N , mientras que la vecindad y la similaridad para realizar la recomendación provienen de múltiples vecindades, creadas a partir de múltiples similaridades y sus *Modelos de Datos*.

Sin la modificación del algoritmo, propuesta en el presente trabajo, un estudiante no podría tener una recomendación por semestre. En su lugar, tendría una recomendación de todo el universo de materias de la carrera, basado en las materias que ha tomado todo el universo de Estudiantes.

5.2. Pruebas realizadas

Para realizar las pruebas se usaron los registros de materias aprobadas de estudiantes de desde el primer término del año 2007 hasta el segundo término del año 2010 de la carrera de Ingeniería en Electrónica y Telecomunicaciones ofertada por la FIEC (Facultad de Ingeniería en Electricidad y Computación), ESPOL.

El motivo por el cual se usó esta carrera se debe a que es la que posee el mayor número de estudiantes registrados en la FIEC, esto es fundamental para el recomendador, ya que la calidad de las recomendaciones está en gran parte determinada por la cantidad y la calidad de los datos usados para realizarlas [19]. Además, la ausencia de especializaciones en la carrera evita que el grupo de estudiantes registrados en ella se dividan en subgrupos, evadiendo así validaciones complejas que introduzcan cambios que pueden afectar los resultados

finales de las recomendaciones.

En cuanto a las pruebas, se basaron en recomendar seis materias a cada uno de los estudiantes de la carrera desde su segundo hasta máximo su octavo semestre, ya que 8 es el número máximo de semestres en el período antes especificado. Para el primer semestre no se recomiendan materias porque no se conocen datos previos de materias aprobadas. Para las recomendaciones de cada semestre los estudiantes que forman parte de ellas, tanto para recibir la recomendación como para formar parte de las vecindades, deben tener aprobado el semestre que se recomienda, para que los resultados de las recomendaciones puedan medir su acierto con decisiones reales.

El método para calcular un valor de acierto para el recomendador con respecto a un estudiante, se define por la ecuación (1):

$$Va=Ma/Mt \quad (1)$$

Donde Ma es el número de materias recomendadas que el estudiante tomo en la realidad y Mt es el número total de materias que el estudiante registro en el semestre recomendado. El valor de acierto final del recomendador es el promedio de todos los valores de acierto calculados de los estudiantes por semestre.

Para realizar las pruebas se usaron cuatro tipos de configuración en el recomendador. Dos variaciones de métrica de similaridad: *Tanimoto*, *Loglikelihood*; y dos de filtrado de vecindad: *Nearest-N*, *Threshold*. La manera en la cual se graficaron los resultados fue mediante la variación de los valores de *Threshold* y *Nearest-N*.

6. Análisis de Resultados

En las pruebas realizadas usando el tipo de vecindad *Nearest-N* se obtuvieron tendencias similares en las curvas de todos los semestres. En la Figura 2 se muestran los resultados obtenidos en el quinto semestre. El eje vertical representa la probabilidad de acierto, mientras que el eje horizontal representa el tamaño de la vecindad. En este tipo de vecindad las curvas de ambas métricas (*Tanimoto* y *Loglikelihood*) crecen de manera exponencial conforme la vecindad se agranda. Esto sucede debido a que en una vecindad de tamaño mayor igual a 4 las recomendaciones se basan en un número de estudiantes no tan reducido, por lo que se cuenta con una cantidad suficiente de ítems a analizar para la recomendación. Los valores óptimos de acierto para todos los semestres se alcanzaron con el tamaño de vecindad 64 y la métrica de similaridad *Loglikelihood* con un valor promedio de 0.83. Cercano a este valor *Tanimoto* alcanzó un acierto promedio de 0.81 para un tamaño de vecindad igual a 32.

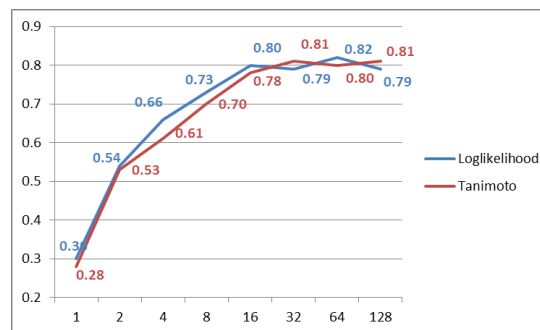


Figura 2. Probabilidad de acierto en los diferentes tamaños de vecindad por *Nearest-N* en el Semestre 5.

El comportamiento de ambas métricas también es similar para el tipo de vecindad *Threshold*. En la Figura 3 se muestran los resultados obtenidos en el quinto semestre. El eje vertical representa la probabilidad de acierto, mientras que el eje horizontal representa el valor del umbral. El incremento del valor del umbral implica que deben tener un historial académico muy similar al del estudiante al que se le realiza la recomendación, esto provoca que el número de estudiantes en la vecindad sea cada vez menor. Como se observa en la Figura 3 conforme el valor de umbral se vuelve más estricto, es decir tiende a 0.9, el valor de acierto obtenido es menor.

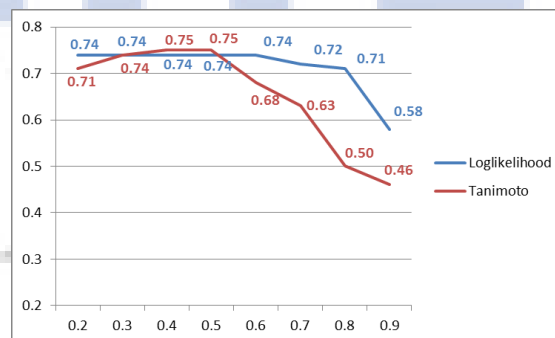


Figura 3. Probabilidad de acierto en los diferentes tamaños de vecindad por *Threshold* en el Semestre 5.

De manera general se observó que a medida que el número de semestre aumenta las curvas empiezan a decrecer de manera más temprana, debido a que a medida que avanzamos en el flujo de una carrera se incrementa el número opciones a tomar en el nuevo semestre. Para este tipo de vecindad *Loglikelihood* demuestro un mejor desempeño con un valor promedio de acierto de 0.79 con un umbral de 0.4 para todos los semestres. *Tanimoto* también obtuvo un valor de acierto promedio de 0.79, sin embargo presentó diferencias de valor considerables entre sus valores máximos y mínimos en cada semestre.

Para los dos tipos de vecindad la métrica *Loglikelihood* alcanzo los valores más altos de acierto. En la Figura 4 se muestran los valores de aciertos que esta métrica obtuvo para cada uno de los semestres

tanto con el tipo de vecindad *Nearest-N* con un tamaño de 64, como con *Threshold* con un umbral de 0.4. El eje vertical representa los valores de acierto alcanzados, mientras que el eje horizontal representa los semestres. Con el tipo de vecindad *Nearest-N* los valores de acierto en las recomendaciones fueron similares para todos los semestres, manteniendo una tendencia más estable alrededor de 0.83, mientras que con el tipo de vecindad *Threshold*, hubo una mayor variación en los valores de acierto alrededor de 0.77.

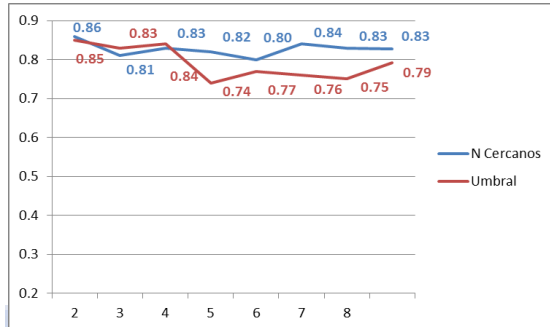


Figura 4. Probabilidad de acierto con la métrica *Loglikelihood* usando los tipos de vecindad *Nearest-N* y *Threshold* del segundo al octavo semestre.

En la Figura 5 se muestra los valores de aciertos que la métrica *Tanimoto* obtuvo para cada uno de los semestres, tanto con el tipo de vecindad *Nearest-N* con un tamaño de 32, como con *Threshold* con un umbral de 0.3. El eje vertical representa los valores de acierto alcanzados, mientras que el eje horizontal representa los semestres. *Tanimoto* presenta una disminución en los valores de acierto conforme avanza en los semestres. A pesar de que estas variaciones son pequeñas contrastan con *Loglikelihood* que mantiene valores estables para todos los semestres. Por este motivo se determinó a *Loglikelihood* con vecindad de tipo *Nearest-N* de tamaño 64 como la combinación óptima.

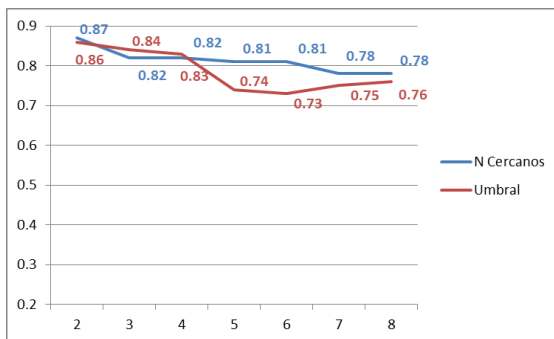


Figura 5. Probabilidad de acierto con la métrica de *Tanimoto* usando los tipos de vecindad *Nearest-N* y *Threshold* del segundo al octavo semestre.

7. Conclusiones

El presente artículo muestra la adaptación de un recomendador basado en el usuario en uno basado en el historial académico del usuario. Las pruebas comprobaron la factibilidad de adaptación presentando niveles adecuados de confiabilidad en un entorno académico, en el caso específico de la carrera de Ingeniería en Electrónica y Telecomunicaciones de la FIEC - ESPOL.

Las pruebas demostraron que para alcanzar la mayor probabilidad de aciertos, de alrededor de 83% para *Loglikelihood* y 81% para *Tanimoto*, el recomendador debe utilizar el tipo de vecindad *Nearest-N* con un tamaño de vecindad de 64 y 32 respectivamente.

Con respecto al tipo de vecindad, el recomendador no sostuvo buenos valores con el uso de vecindad *Threshold*, se notó que al aumentar el valor de umbral el grado de similitud es muy estricto, por lo que no hay muchos estudiantes en los cuales basar la comparación. Con el tipo de vecindad *Nearest-N* se logra ajustar de manera más eficiente la cantidad de estudiantes con similitud alta y mediana que formaran parte de la *Vecindad por Historial*. Con tamaños de vecindad mayores a 64 empiezan a aparecer caídas en los valores de acierto en algunos semestres, a pesar de que las caídas no son de gran magnitud, se concluye que con un tamaño de vecindad grande ingresan estudiantes que introducen “ruido” en las recomendaciones.

Debido a la gran cantidad de información que el recomendador maneja, los tiempos de respuesta en la recomendación para un estudiante estuvieron alrededor de 10 segundos. Si se desea poner en producción el recomendador de materias se deben procesar las recomendaciones de los estudiantes de manera previa al registro de un nuevo semestre. Esta opción es factible debido a que en el escenario de registros de ESPOL los datos en los que se basan las recomendaciones se actualizan solamente al finalizar un semestre. Por el contrario, si los datos tuvieran actualizaciones constantes el recomendador debe ponerse en producción en un ambiente distribuido.

Aproximadamente el 20% de las materias que un estudiante aprueba durante un semestre difiere de las materias que los estudiantes de su vecindad han aprobado. Este porcentaje representa la falta de acierto del recomendador, es decir materias recomendadas que el estudiante objetivo no tomó en ese semestre. Sin embargo, las materias no acertadas son aptas para ser escogidas por el estudiante ya que previamente sus pre requisitos han sido verificados. Además, un gran porcentaje de dichas recomendaciones no acertadas son tomadas por los estudiantes el siguiente semestre al que se le hace recomendación.

En términos generales e independientemente de la configuración que se le dé al recomendador de materias

tanto en su métrica de similaridad como en su filtro de vecindad, se alcanzaron resultados óptimos de recomendación. Consecuentemente, concluimos que la adaptación del recomendador basado en el usuario al recomendador basado en el historial académico del usuario resultó ser exitoso para la evaluación del sexto semestre de la carrera de Ingeniería en Electrónica y Telecomunicaciones.

Como trabajo a futuro se debe considerar añadir al proceso de recomendación otros factores de criterio de elección de materias como: las notas de los estudiantes, valoraciones de los profesores en el CENACAD, etc. Además de contemplar la posibilidad de añadir al recomendador características de los recomendadores basados en el ítem para que el grupo de materias recomendadas sea una combinación más eficiente para el estudiante.

10. Bibliografía

- [1] Miguel-Angel Sicilia, Elena Garcia-Barriocanal, Salvador Sanchez-Alonso, Cristian Cechinel, "Exploring user-based recommender results in large learning object repositories: the case of MERLOT, *Procedia Computer Science, Volume 1, Issue 2, Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning*" (RecSysTEL 2010), *Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)*, 2010, Pages 2859-2864.
- [2] Diz Monje, Cristina y Guzmán Garate, José y Sánchez García, Jesús (2008) "Catalogador automático de textos y recomendador de artículos del portal de revistas electrónicas de la biblioteca UCM". <http://eprints.ucm.es/9887/>
- [3] Lu, J.: Personalized E-learning Material "Recommender System. In *Proceedings of International Conference on Information Technology for Application*" (2004) 374–379.
- [4] Owen S., Robin A. (2009). "Mahout in Action". Greenwich: Manning Publications Co.
- [5] E.J. Castellano and L. Martínez. "A web-decision support system based on collaborative Filtering for academic orientation. case study of the spanish secondary school". *Journal of Universal Computer Science*, 15(14):2786–2807, 2009.
- [6] Lu, J.: "Personalized E-learning Material Recommender System. In *Proceedings of International Conference on Information Technology for Application*" (2004) 374–379.
- [7] J. Sandvig and R. Burke. *Aacorn: A "CBR recommender for academic advising"*. Technical Report TR05-015, DePaul University, 2005.
- [8] Smyth, B.: Case-based Recommendation. In: Brusilovsky, P., Kobsa, A., Nejdl, W.(eds.): "The Adaptive Web: Methods and Strategies of Web Personalization, *Lecture Notes in Computer Science*", Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007).
- [9] N. Bendakir and E. Aimeur. "Using association rules for course recommendation. In *Proceedings of the AAAI Workshop on Educational Data Mining*", pages 31–40, July 16–17 2006.
- [10] R. Farzan and P. Brusilovsky. Social navigation support in a course recommender system. In *Proceedings of the 4th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, pages 91–100, June 21–23 2006.