

Un Sistema de Recomendación de Asignaturas Multi-Criterio con Optimización Genética

Aurora Esteban, Amelia Zafra, Cristóbal Romero
Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Córdoba, España

Resumen—Este artículo propone un Sistema de recomendación (SR) híbrido multi-criterio para resolver el problema de recomendación de asignaturas en los estudios universitarios. Por un lado, utiliza Filtrado Colaborativo (FC) con la información del estudiante y por otro lado, Filtrado Basado en Contenido (FBC) con la información de las asignaturas. Para determinar los factores que resultan más relevantes en la recomendación, así como para optimizar la configuración del SR híbrido en lo referente a las medidas de distancia a considerar y el tamaño del vecindario, se ha diseñado un Algoritmo Genético (AG) que optimiza de forma automática el ajuste de todos los parámetros. Un estudio experimental con 2500 valoraciones reales de los estudiantes del Grado en Ingeniería Informática de la Universidad de Córdoba, demuestra los excelentes resultados obtenidos por el SR propuesto y la importancia de utilizar un modelo híbrido donde los diferentes criterios utilizados sean ponderados de acuerdo a su relevancia.

Index Terms—Sistema de recomendación, Filtrado colaborativo, Filtrado basado en contenido, Algoritmo genético

I. INTRODUCCIÓN

En la era de la información que recientemente se vive, los Sistemas de Recomendación (SR) se han consolidado en diversos campos, entre los que se encuentran los entornos educativos. Específicamente, la recomendación de asignaturas o cursos se ha afianzado como una interesante y creciente línea de investigación dentro de la minería de datos aplicada a la educación [1]. En el caso concreto de la recomendación de asignaturas en los estudios universitarios, su importancia viene dada por el hecho de que los estudios universitarios tienen un número variable de asignaturas optativas entre las que los estudiantes deben elegir para completar los créditos de la titulación que están cursando. Normalmente, los estudiantes no cuentan con información suficiente para hacer esta elección, y recurren a compañeros que la hayan cursado para conocer su opinión. En este contexto, los SR se presentan como una herramienta esencial, capaz de ofrecerles asignaturas relevantes en base a sus preferencias individuales, sus intereses, sus necesidades o su rendimiento [2].

Aunque existen algunos estudios que trabajan con enfoques de SR híbridos [3], [4] o con múltiples criterios [5], [6], no incluyen todos los criterios aquí considerados, ni se

centran en estudiar la influencia que los diferentes factores tienen en el proceso de recomendación.

En este trabajo se presenta un SR híbrido multi-criterio, que combina una técnica de Filtrado Colaborativo (FC) utilizando información del estudiante (considerando como criterios, sus valoraciones, sus calificaciones y la especialidad cursada dentro de la carrera), con una técnica de Filtrado basado en Contenido (FBC) utilizando información de las asignaturas (considerando como criterios, sus competencias, profesores, contenidos teóricos o prácticos y área de conocimiento). Para determinar la importancia de cada criterio se propone un Algoritmo Genético (AG) que optimice los pesos asignados a cada uno de estos criterios en el SR. Así mismo, otros parámetros del SR como las métricas de similitud utilizadas, o el número de vecinos también serán optimizadas simultáneamente. En la metodología propuesta, el SR es configurado con los parámetros optimizados por el AG, para finalmente, realizar las recomendaciones finales a los estudiantes.

Con el fin de garantizar una evaluación rigurosa del SR, se va a utilizar un conjunto de datos proveniente del Grado en Ingeniería Informática de la Universidad de Córdoba (España), que incluye valoraciones y calificaciones de estudiantes reales. Así mismo, en el proceso de evaluación se utilizará validación cruzada manteniendo el equilibrio entre las particiones de datos.

El resto del trabajo se organiza como sigue. En la sección II se hace un repaso de los trabajos previos. En la sección III se especifica la metodología propuesta, describiendo tanto la información, como el SR híbrido diseñado. En la sección IV se explica el AG utilizado para la optimización. En la sección V se describe el estudio experimental realizado, tanto el estudio de la relevancia de los criterios, como la comparativa con otros modelos. Finalmente, la sección VI presenta las conclusiones obtenidas y el trabajo futuro.

II. TRABAJOS RELACIONADOS

En los últimos años los SR multi-criterio han sido ampliamente aplicados a la recomendación de asignaturas o cursos. S. Spiegel [7] explora una de las primeras aplicaciones de factorización de matriz multi-criterio para la predicción de valoraciones de asignaturas. Más adelante, Vialardi et al. [5] propusieron técnicas multi-criterio para la predicción de calificaciones de estudiantes abordadas



como un problema de clasificación, y Parameswaran et al. [8] exploraron la aplicación de restricciones a la recomendación multi-criterio. Así mismo, se ha explorado la aplicación de otras técnicas, como propuestas basadas en ontologías [9], [10], redes neuronales [11], y algoritmos bio-inspirados como colonias de hormigas [12] o sistemas inmunes artificiales [3]. La mayoría de estas propuestas sólo están basados en las valoraciones de los estudiantes. Desde otra perspectiva, la importancia del momento en el que se cursan las asignaturas ha sido estudiada usando Cadenas de Markov en base a las calificaciones [13], así como aplicando multi-criterio [14]. Recientemente, se ha estudiado la relevancia de las competencias que las asignaturas ofrecen [6] o la aplicación de análisis semántico [15].

Se puede resumir que la mayoría de los enfoques se centran principalmente en el rendimiento de los estudiantes, y no usan criterios adicionales. Además, los sistemas que utilizan más criterios, no hacen ningún estudio para determinar la influencia de cada uno de ellos en la calidad de las recomendaciones, mostrando así los beneficios que aporta nuestra propuesta.

III. PROPUESTA DE SISTEMA DE RECOMENDACIÓN HÍBRIDO

En esta sección se describe la metodología propuesta, especificando desde la recolección y preparación de los datos, hasta el desarrollo de la propuesta planteada.

A. Descripción y preparación de los datos

Este trabajo se ha desarrollado utilizando los datos recogidos del Grado en Ingeniería Informática de la Universidad de Córdoba (España). Estos datos son relativos tanto a los estudiantes, como a las asignaturas.

1) *Información de estudiante*: la información de los estudiantes (figura 1) se ha obtenido por medio de encuestas realizadas durante tres cursos académicos (de 2016 a 2018) a los alumnos de los últimos cursos, recogiendo un total de 2500 valoraciones de 63 asignaturas incluidas en el plan de estudios. La información que se ha considerado es:

- Una valoración de la satisfacción general del estudiante en cada asignatura, representada como un valor numérico entero entre 1 y 5. El valor máximo correspondería a un 5, y el mínimo a un 1.
- La calificación obtenida por el estudiante para cada asignatura, representada con un valor numérico real, en el rango $[0, 10]$.
- La especialidad seleccionada por el estudiante. Concretamente, el grado en Ingeniería Informática que se estudia, ofrece tres especialidades: computación, computadores e ingeniería del software, representados con un identificador numérico de 1 a 3.

2) *Información de la asignatura*: la información de las 63 asignaturas consideradas en el plan de estudios (figura

	A1	A2	A3	...	A62	A63	A1	A2	A3	...	A62	A63	
E_i	5	2		...	4		8.5	6.3		...	9		2
	Valoraciones						Calificaciones						Espec.

Figura 1. Información del estudiante

2), se ha obtenido de la página oficial del grado dentro de la Universidad¹. La información que se ha considerado es:

- Los profesores que imparten docencia en cada asignatura, representados con un vector binario, indicando con el valor 1 si el profesor ha impartido la asignatura y con el valor 0 si no la ha impartido.
- Las competencias que se deben adquirir al realizar la asignatura, representadas con un vector binario, indicando con el valor 1 que la asignatura proporciona la competencia, y con el valor 0 que no la proporciona.
- El área de conocimiento al que la asignatura pertenece, representada con un identificador numérico de 1 a 8, ya que en el grado estudiado se contemplan ocho áreas de conocimiento diferentes.
- Los contenidos teóricos y prácticos de la asignatura, representados con un vector de frecuencias de las palabras claves. Las palabras claves son obtenidas a partir del procesamiento automático de textos aplicado a la guía docente de cada asignatura.

	P1	P2	P3	...	P57	P58	Cm1	Cm2	Cm3	...	Cm52	Cm53		
A_i	1	0	1	...	0	1	0	1	1	...	0	1	6	
	Profesores						Competencias						Área conóc.	Contenidos

Figura 2. Información de la asignatura

B. Sistema de Recomendación

El SR híbrido multi-criterio propuesto en este trabajo combina las valoraciones obtenidas por un sistema basado en FC (especificado en la sección III-B1 y las obtenidas por un sistema de FBC (especificado en la sección III-B2). De este modo, la estimación de preferencia p de un estudiante i sobre una asignatura j , se calcula como la combinación lineal de las preferencias dadas por el FC ($FC_{i,j}$) y el FBC ($FBC_{i,j}$):

$$p_{i,j} = \alpha \cdot FC_{i,j} + \beta \cdot FBC_{i,j} \quad (1)$$

Con $\alpha + \beta = 1$

Tanto en el sistema FC, como en el FBC se representan las valoraciones estimadas de cada asignatura en un rango de $[1, 5]$, por lo que la estimación final también estará en dicho rango.

Se puede apreciar en la ecuación 1, que los parámetros que determinan la relevancia que se le da a la valoración obtenida por cada uno de los sistemas considerados son los pesos α y β , que deben ser configurados.

¹<http://www.uco.es/eps/node/619>

En los siguientes apartados se especifican las características del sistema de FC y del sistema de FBC que se han diseñado.

1) *Filtrado Colaborativo utilizando la información del estudiante*: en el sistema de FC desarrollado, la valoración estimada de una asignatura para un estudiante se obtiene a partir de las valoraciones que ésta ha recibido por parte del resto de estudiantes con un perfil similar.

El aspecto multi-criterio es introducido en el cálculo de la similitud entre estudiantes: para cada par de estudiantes i y j , la similitud, s , agrega tres medidas de similitud que están en el rango $[0, 1]$: por un lado, calcula la similitud según sus valoraciones $V_{i,j}$ y sus calificaciones $C_{i,j}$, por otro lado, comprueba si la especialidad que están cursando ambos estudiantes coincide ($E_{i,j}$). Finalmente, los tres criterios se agregan en una combinación lineal:

$$s_{i,j} = \alpha \cdot V_{i,j} + \beta \cdot C_{i,j} + \gamma \cdot E_{i,j} \quad (2)$$

$$\text{Con } \alpha + \beta + \gamma = 1$$

Por tanto, será necesario configurar las métricas utilizadas para valoraciones y calificaciones, el vecindario, y los pesos α , β y γ de los criterios considerados.

2) *Filtrado basado en Contenido utilizando la información de la asignatura*: en el sistema de FBC desarrollado, la valoración estimada de una asignatura para un estudiante se obtiene a partir de la valoración que él mismo ha hecho de las asignaturas más similares a ésta.

El aspecto multi-criterio es introducido en el cálculo de la similitud entre asignaturas: para cada par de asignaturas i y j , la similitud (s) es una combinación de cuatro medidas que están en el rango $[0, 1]$: por un lado se calcula cuántos profesores y competencias tienen en común, $P_{i,j}$ y $Cm_{i,j}$, respectivamente. Por otro lado, se comprueba si el área de conocimiento coincide $A_{i,j}$, y finalmente, se obtiene la similitud según los contenidos aplicando minería de textos $Cn_{i,j}$, siguiendo el procedimiento especificado más adelante. Finalmente, se calcula la similitud final mediante una combinación lineal:

$$s_{i,j} = \alpha \cdot P_{i,j} + \beta \cdot Cm_{i,j} + \gamma \cdot A_{i,j} + \delta \cdot Cn_{i,j} \quad (3)$$

$$\text{Con } \alpha + \beta + \gamma + \delta = 1$$

Por tanto, para aplicar FBC será necesario determinar las métricas usadas para $P_{i,j}$ y $Cm_{i,j}$ y la relevancia de cada criterio por medio de los pesos α , β , γ y δ .

La similitud según los contenidos de la asignatura es obtenida a partir de la información descrita en su guía docente mediante minería de textos. El proceso compara los contenidos de dos asignaturas, para finalmente dar un valor de similitud entre 0 y 1:

1) Indexación del apartado *Contenidos* de las guías docentes de las asignaturas: este proceso se lleva a cabo con un analizador sintáctico y con un conjunto de *stop-words* específicas al dominio. Como resultado

se obtiene para cada documento un conjunto de *tokens* junto a sus frecuencias de aparición.

- 2) Para cada par de asignaturas i y j , se crea un conjunto B que es la unión de sus *tokens*, y para cada asignatura, se crea un vector \vec{i} y \vec{j} que tendrá una posición por cada *token* en B indicando la frecuencia de aparición del mismo. Finalmente, se aplica la norma l_1 sobre cada vector para obtener las frecuencias relativas.
- 3) Finalmente, a los 2 vectores de frecuencias se les aplica la similitud del coseno, obteniendo el valor $Cn_{i,j}$ que se integra en la ecuación parametrizable descrita en 3:

$$\cos(\theta) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \cdot \|\vec{j}\|} = \frac{\sum_{k=1}^n i_k j_k}{\sqrt{\sum_{k=1}^n i_k^2} \sqrt{\sum_{k=1}^n j_k^2}} \quad (4)$$

IV. OPTIMIZACIÓN DEL SISTEMA DE RECOMENDACIÓN MEDIANTE ALGORITMOS GENÉTICOS

El SR que aquí se presenta, descrito en la sección III-B, cuenta con múltiples criterios que deben ser ponderados con un peso que indique la relevancia que va a tener cada uno de ellos, así como con la configuración de diferentes métricas de similitud para cada uno de ellos, y del tamaño del vecindario a utilizar. Por ello, se ha decidido diseñar un AG que automáticamente descubra una configuración óptima de todos los parámetros del SR y permita obtener unas recomendaciones lo más satisfactorias posibles para los estudiantes.

A continuación, se especifica las características del AG que se ha diseñado y que sigue un esquema generacional con elitismo.

A. Representación de los individuos

Los individuos de la población se codifican mediante un cromosoma entero compuesto de 14 genes:

- Los dos primeros genes son los pesos que se van a utilizar para combinar las recomendaciones obtenidas por el FC utilizando la información del estudiante y por el FBC utilizando la información de la asignatura. Concretamente, serían los valores α y β definidos en la ecuación 1.
- Los tres siguientes genes se corresponden con los pesos que se asignan a los tres criterios que se han considerado en el FC utilizando la información del estudiante. Concretamente, serían los valores de α , β y γ definidos en la ecuación 2.
- Los cuatro siguientes genes codifican los pesos que son asignados a los cuatro criterios utilizados en el FBC que emplea la información de la asignatura. Concretamente, serían los valores de α , β , γ y δ definidos en la ecuación 3.
- El siguiente gen es el número de vecinos tenidos en cuenta para realizar las estimaciones.
- Los dos siguientes genes son las métricas utilizadas para calcular la similitud según las valoraciones y las



calificaciones respectivamente. Sus valores son tratados como categóricos, y se corresponderán con las métricas contempladas para obtener estas similitudes.

- Los dos últimos genes se corresponden con las métricas utilizadas para calcular la similitud según los profesores y las competencias. Al igual que el grupo anterior, tomarán valores categóricos que se corresponderán con las métricas contempladas para obtener estas similitudes.

B. Operadores genéticos

1) *El operador de cruce*: es un cruce uniforme que selecciona gen a gen de qué padre toma la información que pasará al hijo. La particularidad de este operador es que el conjunto de genes: (1,2), (3,4,5) y (6,7,8,9) son tratados en bloque, de forma que cada bloque completo es seleccionado del mismo padre para que lo herede uno de los hijos.

2) *El operador de mutación*: es una mutación uniforme, donde para cada individuo que va a ser mutado, se selecciona cada gen con una cierta probabilidad y se cambia su valor por otro valor aleatorio dentro del rango de valores definidos para ese gen.

3) *Optimizador local*: este operador, similar a los empleados en los algoritmos meméticos [16], consiste en realizar una pequeña modificación en los genes que representan las medidas de similitud y el número de vecinos de un mismo individuo. Si el cambio realizado hace que el individuo tenga una mejor función de aptitud, se continúa realizando cambios, hasta un máximo de cinco variaciones. Finalmente, si el individuo obtenido mejora al individuo inicial, lo sustituye en la nueva población, en caso contrario, el nuevo individuo obtenido será eliminado. Debido, al coste computacional de este operador, solamente es aplicado al mejor individuo de la población.

C. La función de aptitud

La función de aptitud o *fitness* que se emplea es la raíz del error cuadrático medio (RMSE) entre las valoraciones reales y las estimadas. Para optimizar los tiempos de cómputo, el fitness se obtendrá utilizando el 80% de los datos como valores conocidos y el restante 20% como datos a estimar para obtener el valor de RMSE.

V. ESTUDIO EXPERIMENTAL

Esta sección describe la experimentación realizada. El SR se ha desarrollado dentro del *framework* Apache Mahout² y el AG utilizando la *librería* JCLEC³. La máquina sobre la que se han ejecutado las pruebas es un ordenador personal con SO Ubuntu 16.04 64-bits, un procesador Intel Core i5-3317U y 12 GB de RAM.

²<https://mahout.apache.org/>

³<http://jclec.sourceforge.net/>

A. Configuración del Algoritmo Genético

Los principales parámetros utilizados en el AG, tras un estudio de diferentes configuraciones, son: tamaño de la población: 200, número de generaciones: 100, y probabilidad de mutación y cruce: 0.5 y 0.9 respectivamente.

Con respecto al rango de valores de cada gen: los 9 primeros genes que representan los pesos de cada criterio considerado, se han definido en el rango [0, 10]. Este rango podría variarse, ya que estos valores son normalizados antes de ser utilizados en las ecuaciones 1, 2 y 3. El gen 10, que representa el número de vecinos, se ha definido en el rango [1, 50]. Los genes 11 y 12 que representan las medidas de similitud de las valoraciones y las calificaciones, se definen en el rango [0, 4] representando las 5 métricas consideradas, la distancia euclidiana, la distancia Manhattan, el coeficiente de correlación de Pearson, el coeficiente de correlación de Spearman y la similitud del coseno. Finalmente, los genes 13 y 14 que representan las medidas de similitud de los profesores y las competencias, se definen en el rango [0, 1] representando las 2 métricas consideradas, el índice Jaccard y la función de verosimilitud logarítmica.

B. Influencia de los diferentes criterios en el Sistema de Recomendación

En esta sección, se evalúa la relevancia de cada uno de los criterios considerados en el SR híbrido presentado, para determinar los factores que pueden resultar más significativos para mejorar las recomendaciones.

En la tabla I se muestra la mejor configuración de pesos encontrada por el AG. En el caso de la configuración del FC que emplea información del estudiante, se puede apreciar que el factor más relevante es la calificación obtenida por los estudiantes (peso de 0.53), seguido por las valoraciones dadas a las asignaturas (peso de 0.25) y la especialidad cursada (peso de 0.22). Estos resultados muestran la relevancia de las calificaciones en la búsqueda

Cuadro I
MEJOR CONFIGURACIÓN DEL SR OBTENIDA POR EL AG

SR Híbrido	
Peso del FC	0.58
Peso del FBC	0.42
FC (Basado en información del estudiante)	
Valoraciones (métrica sim.)	Distancia Manhattan
Calificaciones (métrica sim.)	Distancia Manhattan
Valoraciones (peso)	0.25
Calificaciones (peso)	0.53
Especialidad (peso)	0.22
Tamaño vecindario	12
FBC (Basado en información de la asignatura)	
Profesores (métrica sim.)	Verosimilitud logarítmica
Competencias (métrica sim.)	Índice de Jaccard
Profesores (peso)	0.54
Competencias (peso)	0.00
Área conoc. (peso)	0.00
Contenido (peso)	0.46

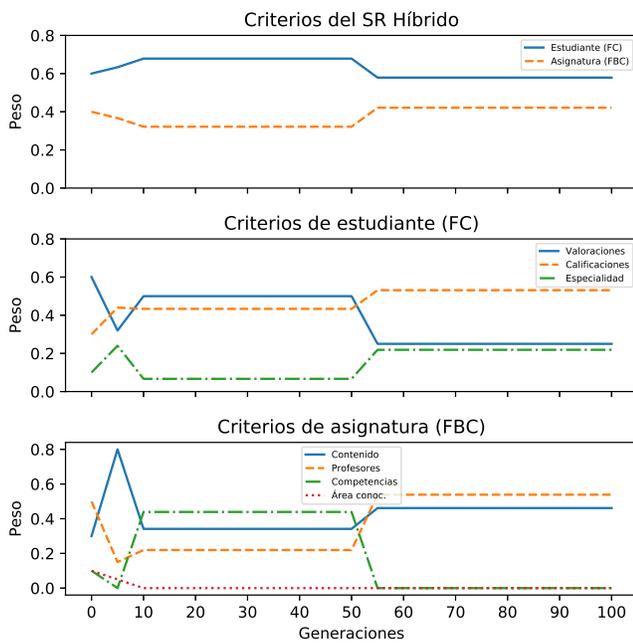


Figura 3. Evolución de los pesos de los criterios en el AG

de estudiantes similares. Así, una asignatura será recomendada a un cierto estudiante si, principalmente, ha sido valorada positivamente por otros estudiantes con unas calificaciones similares a la suya.

Analizando los criterios más relevantes relativos a la información de la asignatura (FBC), se obtiene como factores muy relevantes los profesores que imparten las asignaturas (0.54) y el contenido teórico y práctico de las mismas (0.46); los otros dos criterios (competencias que cubren las asignaturas y el área de conocimiento a la que pertenecen), aparecen como irrelevantes (0.00). Estos resultados, ponen de manifiesto la importancia de los profesores, siendo un factor tan importante como el contenido propio de la asignatura para determinar cómo de similares se perciben dos asignaturas. De esta manera, una asignatura será recomendada a un cierto estudiante si tiene un contenido similar y/o ha sido impartida por el mismo profesorado que otra asignatura que le haya interesado en el pasado.

Finalmente, los pesos asignados por el AG a las estimaciones realizadas por los sistemas FC y FBC son ponderados con 0.58 y 0.42 respectivamente, mostrando la importancia de considerar tanto información del estudiante, como de la asignatura en la recomendación final.

Para estudiar cómo los pesos han ido evolucionando en las sucesivas generaciones del AG, se analiza la evolución del mejor individuo de la población (Figura 3). Se aprecia que los pesos asignados a las recomendaciones realizadas por el FC y por el FBC, comienzan bastante desequilibradas, dando más importancia al FC. Sin embargo, acaban convergiendo en una importancia similar. En el caso de

los pesos asignados a los diferentes criterios considerados en la información del estudiante, se puede apreciar como en las soluciones iniciales el factor de las valoraciones es el más relevante. No obstante, conforme el AG va convergiendo y optimizando el ajuste de parámetros, las calificaciones toman más importancia, pasando valoraciones y especialidad a tener una relevancia similar y menor a las calificaciones. Finalmente, si se estudia la evolución de los pesos asignados por el AG a los diferentes criterios relativos a la información de las asignaturas, se puede apreciar que desde el principio, el área de conocimiento aparece como poco significativa, y es algo que se mantiene hasta el final, mientras que el contenido y los profesores va tomando más relevancia cada vez.

Se puede apreciar que a partir de la generación 60 el peso asignado a los diferentes criterios se mantiene, no siendo mejorados por ninguna otra solución de la población durante el resto de las generaciones. El motivo de que el AG continúe realizando generaciones y optimizando soluciones se debe a que continúa obteniendo mejores soluciones optimizando el resto de parámetros considerados en el AG y configurables en el SR híbrido diseñado, que son las métricas de similitud utilizadas y el tamaño del vecindario.

En la tabla I se muestra las métricas de similitud y el tamaño de vecindario seleccionadas por el AG. En la figura 4, se muestra la evolución del mejor individuo de la población a lo largo de las generaciones para estos valores. En este caso, es interesante resaltar como estos valores van cambiando, a partir de la generación 60, para optimizar las valoraciones proporcionadas por el SR y en consecuencia reducir el RMSE. Se puede apreciar que existe cierta tendencia a que los criterios de valoraciones y calificaciones utilicen la misma métrica. Además, si esta métrica es la medida de similitud del coseno, se necesita

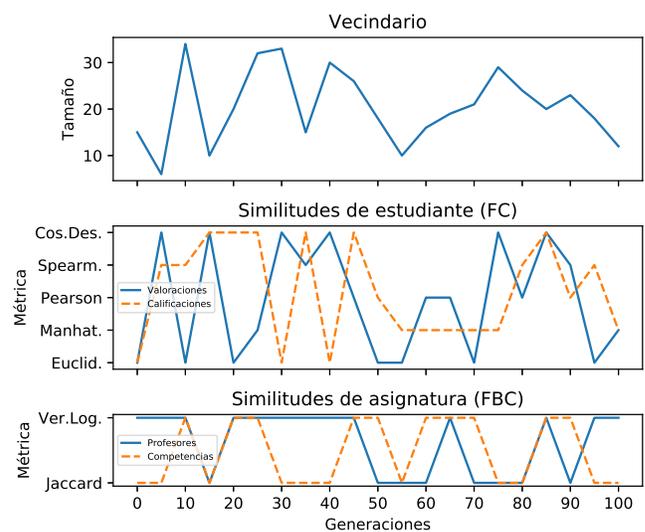


Figura 4. Evolución de las métricas y el vecindario en el AG



un mayor número de vecinos, mientras que para el resto de casos tiende a mantenerse entre 10 y 20.

C. Comparativa con otros modelos

Para concluir el estudio experimental, se compara el rendimiento del SR híbrido frente al uso de los modelos: FC multi-criterio y FBC multi-criterio con la misma configuración de pesos establecida por el AG, pero sin combinar sus resultados. También se incluye en la comparativa el uso de FC mono-criterio empleando solamente el criterio de las valoraciones (el criterio más ampliamente utilizado en los SR) y FBC mono-criterio empleando solamente la información del criterio del profesorado (el más representativo, de acuerdo a los pesos asignados).

Para llevar a cabo este estudio, se ha ejecutado una evaluación del SR utilizando validación cruzada con 10 *fold* con muestreo estratificado. Las métricas estudiadas son el RMSE, relacionado con cuánto difiere una estimación de la valoración real; la ganancia nDCG (*normalized Discounted Cumulative Gain*), relacionada con la capacidad del SR de ofrecer recomendaciones relevantes; el alcance, que determina el porcentaje de usuarios para los que se pueden obtener recomendaciones; y el tiempo de ejecución que necesita el SR para entrenar el modelo y ofrecer una recomendación.

Los resultados finales pueden verse en la tabla II. El enfoque híbrido obtiene mejores resultados en todas las métricas estudiadas. Aunque al usar más información, tarda algo más que los otros modelos en ejecutarse. De este modo, queda demostrada la importancia de utilizar múltiples criterios en el SR, asignándoles además pesos concretos, así como la de realizar una hibridación de diferentes técnicas.

Cuadro II
COMPARATIVA ENTRE MODELOS DE SR

Método	RMSE	nDCG	Alcance(%)	Tiempo(s)
SR Híbrido	1.056	0.811	100.00	8.23
FC ¹	1.167	0.806	96.48	7.53
FBC ¹	1.201	0.214	99.36	6.86
FC ²	1.233	0.798	96.48	5.98
FBC ²	2.530	0.284	99.36	5.77

¹ multi-criterio, ² mono-criterio

VI. CONCLUSIONES

En este trabajo se ha desarrollado un SR aplicado a la recomendación de asignaturas híbrido y multi-criterio. Así mismo, se ha implementado un AG que realiza un ajuste de todos los parámetros utilizados en el SR propuesto para determinar las configuraciones óptimas para el SR. Las pruebas realizadas muestran que considerar varios criterios proporciona mejores resultados, pero que la relevancia de cada uno de ellos debe ser estudiada, ya que no todos los factores resultan igual de relevantes. Además, la consideración de un sistema híbrido, que combina tanto filtrado colaborativo, como basado en contenido, también optimiza los resultados alcanzados.

En un futuro se pretende ampliar las pruebas a otras titulaciones, de forma que se pueda comprobar si se obtienen los mismos resultados, y bajo qué circunstancias se pueden generalizar las conclusiones.

VII. AGRADECIMIENTO

Este trabajo está financiado por el proyecto de investigación TIN2017-83445-P del Ministerio de Economía y Competitividad y el Fondo Europeo de Desarrollo Regional.

REFERENCIAS

- [1] C. Romero and S. Ventura. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1):12–27, 2013.
- [2] T.N. Huynh-Lv, N. Huu-Hoa, and T.N. Nguyen. Methods for building course recommendation systems. In *Proceedings of the 8th International Conference on Knowledge and Systems Engineering*, pages 163–168, 2016.
- [3] P.-C. Chang, C.-H. Lin, and M.-H. Chen. A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. *Algorithms*, 9(3):1–18, 2016.
- [4] C. Kim, N. Choi, Y. Heo, and J. Sin. On the development of a course recommender system: A hybrid filtering approach. *Entrée Journal of Information Technology*, 14(2):71–82, 2015.
- [5] C. Vialardi, J. Chue, J.P. Peche, G. Alvarado, B. Vinatea, J. Estrella, and A. Ortigosa. A data mining approach to guide students through the enrollment process based on academic performance. *User Modeling and User-Adapted Interaction*, 21(1-2):217–248, 2011.
- [6] J.W. Han, J.C Jo, H.S. Ji, and H.S. Lim. A collaborative recommender system for learning courses considering the relevance of a learner's learning skills. *Networks Software Tools and Applications*, 19(4):2273–2284, 2016.
- [7] S. Spiegel. *A Hybrid Approach to Recommender Systems based on Matrix Factorization*. Thesis, Tech. University Berlin, 2009.
- [8] A. Parameswaran, P. Venetis, and H. Garcia-Molina. Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems*, 29(4):1–33, 2011.
- [9] C.Y. Huang, R.C. Chen, and L.S. Chen. Course-recommender system based on ontology. In *Proceedings of 12th International Conference on Machine Learning and Cybernetics*, pages 1168–1173, 2013.
- [10] L. Zhuhadar, O. Nasraoui, R. Wyatt, and E. Romero. Multi-model ontology-based hybrid recommender system in e-learning domain. In *Proceedings of the International Joint Conferences on Web Intelligence and Intelligent Agent Technology*, pages 91–95, 2009.
- [11] A.A. Kardan, H. Sadeghi, S.S. Ghidary, and M.R.F. Sani. Prediction of student course selection in online higher education institutes using neural network. 65:1–11, 2013.
- [12] J. Sobacki and J.M. Tomczak. Student courses recommendation using ant colony optimization. In *Proceedings of the 2nd Intelligent Information and Database Systems*, volume 5991 of *Lecture Notes in Artificial Intelligence*, pages 124–133, 2010.
- [13] E. Khorasani, Z. Zhenge, and J. Champaign. A markov chain collaborative filtering model for course enrollment recommendations. In *Proceedings of the 4th International Conference on Big Data*, pages 3484–3490. IEEE, 2016.
- [14] R. Wang. *Sequence-based Approaches to Course Recommender Systems*. Thesis, University of Alberta, 2017.
- [15] H. Ma, X. Wang, J. Hou, and Y. Lu. Course recommendation based on semantic similarity analysis. In *Proceedings of 3rd Ieee International Conference on Control Science and Systems Engineering*, pages 638–641, 2017.
- [16] E. Özcan and C. Başaran. A case study of memetic algorithms for constraint optimization. *Soft Computing*, 13(8-9):871–882, 2009.