

Algoritmos de filtrado colaborativo para recomendación de hilos en foros de cursos MOOC

José Antonio González Martínez, Eduardo Gómez Sánchez, Miguel L. Bote Lorenzo
Grupo de Sistemas Inteligentes y Cooperativos / Educación, Medios, Informática y Cultura
Escuela Técnica Superior de Ingenieros de Telecomunicación. Universidad de Valladolid.
Paseo de Belén 15, 47011, Valladolid, España
jgonmar@gsic.uva.es, edugom@tel.uva.es, migbot@tel.uva.es

Resumen— Este estudio explora los retos y oportunidades del filtrado colaborativo para recomendar hilos en foros de discusión de cursos MOOC. Se consideran métricas clásicas y nuevas para estimar la similitud entre usuarios, se discute la definición de una *ground truth* y se plantea un estudio experimental, llevado a cabo con datos reales, en el que el sistema hace recomendaciones no obvias con la información disponible en cada día del curso.

Palabras clave—sistemas de recomendación, MOOC, foros de discusión, filtrado colaborativo.

I. INTRODUCCIÓN

Los cursos MOOC (o *Massive Open Online Courses*) son una modalidad reciente de aprendizaje en línea que permite potencialmente a cientos de miles de usuarios aprender sobre una gran variedad de materias ofrecidas por las más prestigiosas organizaciones del mundo [1]. Los foros de discusión de los MOOC son centros de actividad muy importantes para los alumnos [2], de los que se sirven para pedir y ofrecer ayuda en la resolución de ejercicios y dudas. También es, en muchos casos, la única vía para relacionarse y colaborar con otros usuarios [3].

Los foros educativos ofrecen a los usuarios funcionalidades únicas para el debate, la resolución de problemas y la colaboración [4]. Sin embargo, un aprendiz también puede encontrarse con problemas causados por la organización y cantidad de información del foro. El alumno puede no encontrar la información relevante [4] o perder mucho tiempo buscándola, lo que puede incluso desmotivarle [5]. Para evitarlo, los foros generalmente incorporan herramientas de búsqueda, pero el usuario a veces no consigue expresar sus intereses para orientar correctamente la consulta [6]. Todos estos problemas son especialmente graves en foros MOOC, donde puede haber miles de usuarios e hilos.

En este contexto, parece deseable pedagógicamente ofrecer funcionalidades en los foros MOOC para proporcionar recomendaciones personalizadas a los alumnos. Por una parte, un sistema de este tipo permitiría resaltar información relevante que de otra manera puede permanecer oculta. Además, podría mejorar la sencillez y la utilidad percibida del foro, lo que evitaría la desmotivación del aprendiz. Por último, las recomendaciones de hilos mejorarían la participación en los foros, ya que los alumnos contribuirán con mayor probabilidad si encuentran información suficientemente interesante. Para ello, una de las técnicas de recomendación más ampliamente

utilizadas es el llamado filtrado colaborativo. Este es un estudio exploratorio de los retos y ventajas de utilizar filtrado colaborativo para recomendar hilos en foros de discusión de cursos MOOC, como paso previo para mejorar su utilidad, fomentando la participación de los alumnos y su motivación, y aumentando la probabilidad de que completen el curso.

El artículo se estructura de la siguiente manera. La sección 2 tratará el estado del arte en sistemas de recomendación, especialmente en el entorno educativo y en foros MOOC. En la sección 3 se explicarán las oportunidades que surgen al aplicar algoritmos CF en foros MOOC. La sección 4 detallará el estudio experimental desarrollado para evaluar la eficacia de algoritmos CF sobre un conjunto de datos real. La sección 5 discute los resultados obtenidos y la sección 6 recoge las conclusiones y trabajo futuro.

II. ESTADO DEL ARTE

En la literatura, los sistemas de recomendación se han clasificado en dos grandes grupos [7]. Por un lado, en los sistemas de recomendación basados en contenido, al usuario se le ofrecen elementos en función de una descripción del elemento y el perfil de los intereses del usuario [8]. Por otro lado, en los sistemas de filtrado colaborativo (*Collaborative Filtering*, CF), se recomiendan objetos que son valorados positivamente por usuarios similares [9]. Los sistemas de recomendación han sido ampliamente utilizados en diferentes dominios, como el comercio electrónico o el entretenimiento (por ejemplo, recomendando productos en Amazon [10] o vídeos de YouTube [11]), y también se han aplicado en el dominio del aprendizaje apoyado por la tecnología [9].

Algunas contribuciones han estudiado sistemas de recomendación en foros de discusión de propósito general (por ejemplo, [12] usando filtrado colaborativo y [13] realizando análisis semántico). Mientras, otros estudios tratan recomendadores específicamente en foros educativos, como Comtella-D [14], un foro educativo con filtrado colaborativo para sugerir hilos a los alumnos. En foros MOOC, se está empezando a investigar sobre sistemas de recomendación. Por ejemplo, [15] y [16] proponen sendos recomendadores de hilos y preguntas en foros MOOC. Por otra parte, [17] presenta un modelo que predice las respuestas aceptadas a preguntas en foros MOOC. Por último, [18] muestra un sistema que recomienda vídeos formativos si detecta confusión en las preguntas publicadas en foros MOOC. Sin embargo, hasta el

momento no se conocen trabajos que utilicen algoritmos CF para recomendar hilos en foros MOOC, cuya aplicación podría dar buenos resultados, como se discutirá a continuación.

III. OPORTUNIDADES DEL FILTRADO COLABORATIVO EN FOROS MOOC

Los sistemas de recomendación CF tienen un potencial de aplicación alto en entornos donde hay muchos elementos y muchas valoraciones por elemento, hay más valoraciones que elementos a recomendar, los usuarios valoran múltiples elementos, estos elementos son homogéneos y hay usuarios con necesidades comunes [19]. Todas estas características se cumplen en general en los foros MOOC. Por ejemplo, el foro del curso 6.002x "*Circuits and Electronics*" de la primavera de 2013, organizado por el *Massachusetts Institute of Technology* (MIT), contenía cerca de 6.000 mensajes en unos 1.400 hilos publicados por alrededor de 1.000 usuarios, que votaron los hilos 2.866 veces e hicieron cerca de 150.000 visitas a hilos. Adicionalmente, se cree [19] que las técnicas CF producen recomendaciones de elementos inesperados (serendipia), lo que podría ser útil en foros MOOC donde puede haber hilos relevantes que pasen desapercibidos para muchos usuarios.

Por el contrario, también es cierto que las técnicas CF son menos adecuadas en dominios en los que la valoración de un elemento no requiere un gusto personal, el interés de los elementos a recomendar no persiste en el tiempo [19] (ya que los hilos pueden perder interés según avanzan los temas de estudio), o la dispersión de la matriz de valoraciones es grande, características también presentes en los foros MOOC. Puesto que aún no se han aplicado técnicas de CF en la recomendación en estos foros, es interesante conocer su comportamiento.

IV. ESTUDIO EXPERIMENTAL: RETOS Y RESULTADOS

Esta sección presenta los resultados del estudio exploratorio realizado para evaluar el comportamiento de algoritmos CF para la recomendación de hilos en foros MOOC. Para poder aplicar estos algoritmos, se tuvieron que resolver una serie de retos asociados, que se explican a continuación.

Los experimentos se realizaron con un conjunto de datos del curso 6.002x "*Circuits and Electronics*", que comenzó el 13 de marzo y finalizó el 20 de junio de 2013. El conjunto de datos incluye información completa sobre el foro (contenidos y metadatos de hilos y mensajes), usuarios, trazas de eventos, etc. El foro de discusión está compuesto por 1.390 hilos generados por 1.079 usuarios y 5.947 mensajes.

A. Diseño del sistema de recomendación de hilos

El primer paso consistió en diseñar el tipo de recomendador a utilizar. Para ello, se decidió utilizar un sistema de recomendación CF para recomendar 3 hilos relevantes a cada usuario. Este número de elementos se eligió porque no es muy grande, es fácilmente asimilable por el usuario, y es posible embeberlos para presentarse al usuario en la misma página web de los foros MOOC. El recomendador sugerirá hilos relevantes, por lo que entra dentro de la categoría de recomendadores cuya tarea es "proponer buenos elementos" [20], ya que el objetivo es mostrar contenido de interés para el usuario, sin que sea necesario que estén ordenados. Se decidió utilizar el algoritmo CF basado en vecindad, por ser

ampliamente utilizado en la literatura [19]. Además, se generarán recomendaciones para cada día del curso y no se recomendarán hilos ya visitados por el usuario, puesto que no tiene sentido recomendar hilos que el usuario ya conoce [20].

B. Selección de la métrica de valoración para los hilos

El segundo reto consistió en determinar cuál sería la métrica de valoración de un hilo. En los foros MOOC los usuarios no puntúan la utilidad de un hilo. Por ello, de entre la información disponible (el número de visitas a un hilo, el número de mensajes publicados, los votos positivos en un hilo o el tiempo de lectura que el usuario le ha dedicado), es necesario seleccionar la métrica adecuada que dé idea de la relevancia de un hilo para un usuario. La información del número de votos positivos y el número de mensajes de un usuario en un hilo se descartó porque era escasa. También porque puede haber usuarios a los que les guste el hilo pero que por otros motivos (pereza, olvido, etc.) no lo voten positivamente. Además, interesa diseñar un recomendador genérico y es posible que la funcionalidad de votos positivos no esté disponible en todas las plataformas MOOC. También se descartó utilizar el tiempo de visita de un usuario en un hilo como indicador implícito de su utilidad, por la complejidad y carga que supone procesar eventos en las trazas para calcular el tiempo que había pasado un usuario en un hilo.

Por lo tanto, se decidió utilizar la información del número de visitas de un usuario en un hilo, que es muy abundante, como indicador de la utilidad de ese hilo. Esta métrica parece adecuada, ya que un hilo visitado muchas veces puede indicar que su contenido es interesante para el usuario, mientras que si solo lo ha visitado una vez, puede sugerir que el usuario lo ha leído, pero es suficientemente relevante como para consultarlo de nuevo. Los datos apoyan esta suposición porque un usuario que ha votado positivamente un hilo, lo ha visitado como media 4.74 veces, mientras que la media de visitas en hilos visitados pero no votados positivamente es de 1.79. A partir del número de visitas, se otorgó un valor de 0 cuando el hilo no tiene visitas; un valor de -1 cuando el hilo sólo tiene 1 visita (ha sido visitado, pero el usuario no lo ha encontrado relevante porque no lo ha vuelto a visitar); y un valor de +1 cuando el hilo ha tenido más de una visita (ha sido visitado en más de una ocasión porque el usuario lo ha encontrado relevante). Para cada día del curso, se generó una matriz de valoraciones con esta información ternaria, en la que las filas eran los usuarios del foro y las columnas los hilos.

C. Selección de la métrica de similitud entre usuarios

Puesto que las técnicas de filtrado colaborativo se basan en el parecido entre vecinos, el siguiente reto fue determinar cómo se calculará la similitud entre usuarios. Existen heurísticas clásicas para evaluar el grado de semejanza entre dos usuarios (por ejemplo, la correlación de Pearson o el vector coseno) [7]. Otras métricas diferentes pueden dar la misma o mejor información y ofrecer un rendimiento superior.

En primer lugar, se podría suponer que un usuario se parece mucho a otro si ha visitado los mismos hilos. En este caso, la similitud se mediría por el número de hilos covisitados. Alternativamente, podríamos asumir que un usuario es más parecido a otro, cuantos más hilos hayan considerado de la misma forma. Esta similitud se podría calcular con el producto

escalar de los vectores de los dos usuarios. Interesa comparar el efecto de usar estas métricas en la precisión del algoritmo de recomendación. En este estudio, se realizarán pruebas con la similitud de vector coseno, por ser una métrica habitual en el ámbito CF; y la similitud por hilos covisitados y la similitud producto escalar porque podrían capturar la similitud entre usuarios que visitan hilos de interés común.

D. Definición de la pseudo ground truth

Para realizar el proceso de evaluación, se necesita disponer de una *ground truth* o valoraciones reales de referencia. Con esta referencia se podría determinar si el sistema ha acertado con sus recomendaciones y evaluar su eficacia. En un foro MOOC, no es inmediato contar con una *ground truth*, ya que los usuarios no puntúan los hilos. Además, tampoco es posible realizar en nuestra experimentación una evaluación en línea en la que usuarios reales utilicen el sistema y valoren explícitamente los elementos recomendados [21] ni utilizar a expertos para valorar la relevancia de los hilos como si fueran usuarios (lo que se conoce como *cross-rating*). Por lo tanto, se hace necesario definir un método indirecto para obtener una estimación de la *ground truth* o *pseudo ground truth* con la información disponible, lo más fiel posible a la realidad.

Para calcular la *pseudo ground truth* se utilizaron todos los datos disponibles sobre las visitas a los hilos al final del curso (20 de junio de 2013), ya que esta información final puede resumir el interés de cada usuario por los hilos. En consonancia con la selección de la métrica de valoración de hilos, un hilo es relevante para un usuario si lo ha visitado más de una vez al final del curso, y no es relevante si ha sido visitado sólo una vez. Un hilo que no ha sido visitado nunca por un usuario no se tiene en cuenta para la evaluación de ese usuario, porque no sabemos si lo consideraría relevante o no si lo hubiera visitado.

E. Proceso de evaluación

Este reto consistió en diseñar el propio proceso de evaluación. Se decidió calcular la media de los aciertos del sistema de recomendación a lo largo de 3 semanas del curso, (del 14 de marzo al 3 de abril de 2013). Se utilizó la precisión media o MAP (*Mean Average Precision*) para determinar la eficacia del algoritmo. Se usó esta métrica por el tipo de recomendador diseñado, en el que los elementos se valoran en una escala ternaria, y solo se muestra una lista limitada de 3 elementos relevantes sin implicar un orden de los mismos [22].

En cada instante de tiempo t y para un determinado usuario, sólo se evalúan los hilos que ya existen en ese momento, que todavía no han sido visitados por el usuario (si no sería una recomendación obvia), pero que se sabe que lo serán en el futuro (de manera que se puede aplicar la *pseudo ground truth* para evaluar la precisión). Además, solo serán evaluados aquellos alumnos que en el instante t tengan al menos 3 hilos relevantes en la *pseudo ground truth*. Así, aseguramos que la precisión máxima posible es 1.

Los resultados obtenidos tras evaluar el algoritmo de filtrado colaborativo a lo largo de los 21 días del periodo de observación se recogen en la Figura 1. Se compara la MAP obtenida por el algoritmo CF con las 3 métricas utilizadas (coseno -COS-, covisitados -CV- y producto escalar -PROD-), además de una heurística *naive*, que utilizaremos como

baseline, en la que se recomiendan a los usuarios los 3 hilos más visitados en ese momento que el alumno todavía no haya visitado. Una heurística similar se utilizó en [16]. Se puede ver que los mejores resultados de la precisión se consiguieron con el algoritmo CF que utiliza la similitud de hilos covisitados. En algunos momentos del curso, la precisión llegó a ser superior al 70% (similar a la obtenida en [16]). La segunda métrica con mejores resultados medios fue la similitud por vector coseno, con resultados muy parecidos a la heurística *baseline*. Los resultados promedios de la Tabla 1, muestran también el mejor comportamiento de la similitud de hilos covisitados.

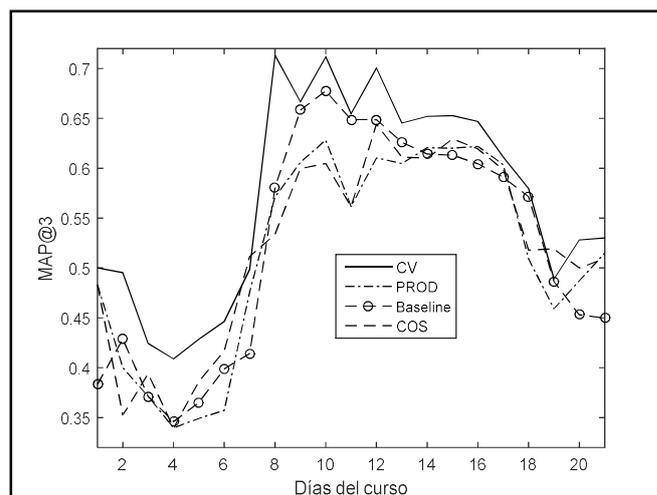


Figura 1 Precisión media a lo largo de 21 días del curso

TABLA 1 MAP@3 promediado los 21 días del periodo de estudio

	CV	COS	Baseline	PROD
MAP@3	57.12%	52.11%	52.05%	51.40%

V. DISCUSIÓN

Los algoritmos CF ofrecen un buen comportamiento en la recomendación de hilos, si se comparan con el *baseline* o con los resultados ofrecidos por otros sistemas de recomendación para foros MOOC (por ejemplo, el 32.3% de MAP@3, según [17]). Las mejores prestaciones se consiguieron utilizando la métrica covisitados. Esta métrica determina que un usuario es más parecido a otro cuantos más hilos hayan visitado en común. Las otras métricas no parecen captar tan fielmente que dos usuarios que visiten hilos en común, tengan patrones de visita y, probablemente, intereses similares en el foro. Por otro lado, una característica diferenciadora de este estudio es que evalúa la calidad de un recomendador que sugiere hilos a lo largo del tiempo con la información disponible en ese momento. Otros estudios ([15], [16]) sólo reflejan una predicción al final del curso, lo que no es muy útil en un contexto real. Se observa un comportamiento variable de la precisión en el tiempo, siendo superior al 50% en el tramo central del periodo estudiado. Es interesante ver que, comparando esto con el calendario del curso¹, las transiciones coinciden con momentos donde se publican ejercicios de esa semana o es la fecha límite para entregarlos. Esta información podría ser explotada para mejorar las recomendaciones.

¹ https://courses.edx.org/c4x/MITx/6.002x/asset/sp2013_6.002x_calendar.pdf

VI. CONCLUSIONES

Este estudio ha explorado las dificultades y el potencial del filtrado colaborativo para la recomendación de hilos en foros MOOC. En el cálculo de la similitud entre usuarios se ha encontrado que las distancias clásicas, como el coseno, en un espacio de valoraciones enormemente disperso, ofrecen peores resultados que otras métricas más sencillas y computacionalmente más eficientes, como el número de elementos covalorados. La definición de una *pseudo ground truth* es otro problema relevante, al carecerse en los foros MOOC de suficientes valoraciones explícitas por parte de los usuarios. Se ha propuesto derivar el interés del usuario por un hilo a partir de su número de visitas, aunque otras opciones serán exploradas en el trabajo futuro. Por último, en los experimentos se ha evaluado la calidad del recomendador actuando en cada instante de tiempo del curso con la información disponible en ese momento, lo que lo acerca a las necesidades de un sistema en producción, en contraste con otros trabajos en los que sólo se evalúa la recomendación al final del curso. Los resultados alcanzados apuntan a la utilidad del filtrado colaborativo para la recomendación de hilos en foros MOOC, y al potencial de considerar información contextual del momento del curso para mejorarla. Esto será explorado en el trabajo futuro, además de la posibilidad de determinar la afinidad entre usuarios a partir de elementos diferentes del MOOC (vídeos o ejercicios, por ejemplo).

AGRADECIMIENTOS

El acceso a los datos utilizados en este trabajo se consiguió gracias a un acuerdo suscrito con la Oficina de Investigación Institucional del *Massachusetts Institute of Technology*. Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad del Gobierno de España (proyecto de investigación TIN2014-53199-C3-2-R) y la Consejería de Educación de la Junta de Castilla y León (proyecto de investigación VA277U14).

REFERENCIAS

- [1] N. Gillani, "Learner communications in massively open online courses," *OxCHEPS Occas. Pap.*, no. 53, 2013.
- [2] D. T. Seaton, Y. Bergner, I. Chuang, P. Mitros, and D. E. Pritchard, "Who Does What in a Massive Open Online Course?," *Int.J.Hum.-Comput.Stud.*, vol. 68, p. 223, 2010.
- [3] D. F. O. Onah, J. Sinclair, and R. Boyatt, "Exploring the use of MOOC discussion forums," 2014.
- [4] F. Abel, I. I. Bittencourt, E. Costa, N. Henze, D. Krause, and J. Vassileva, "Recommendations in online discussion forums for e-learning systems," *Learn. Technol. IEEE Trans.*, vol. 3, no. 2, p. 165, 2010.
- [5] P.-Y. Wang and H.-C. Yang, "Using collaborative filtering to support college students' use of online forum for English learning," *Comput. Educ.*, vol. 59, no. 2, p. 628, 2012.
- [6] A. Sieg, B. Mobasher, and R. Burke, "Inferring user's information context from user profiles and concept hierarchies," in *Classification, Clustering, and Data Mining Applications*, Springer, 2004, p. 563.
- [7] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *Knowl. Data Eng. IEEE Trans.*, vol. 17, no. 6, pp. 734–749, 2005.
- [8] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The adaptive web*, Springer, 2007, p. 325.
- [9] N. Manouselis, H. Drachler, K. Verbert, and E. Duval, *Recommender systems for learning*. Springer, 2013.
- [10] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *Internet Comput. IEEE*, vol. 7, no. 1, p. 76, 2003.
- [11] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, and B. Livingston, "The YouTube video recommendation system," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, p. 293.
- [12] C. Castro-Herrera, J. Cleland-Huang, and B. Mobasher, "A recommender system for dynamically evolving online forums," in *Proceedings of the third ACM conference on Recommender systems*, 2009, p. 213.
- [13] Y.-M. Li, T.-F. Liao, and C.-Y. Lai, "A social recommender mechanism for improving knowledge sharing in online forums," *Inf. Process. Manag.*, vol. 48, no. 5, pp. 978–994, 2012.
- [14] F. Abel, I. I. Bittencourt, N. Henze, D. Krause, and J. Vassileva, "A rule-based recommender system for online discussion forums," in *Adaptive Hypermedia and Adaptive Web-Based Systems*, 2008, p. 12.
- [15] D. Yang, M. Piergallini, I. Howley, and C. Rose, "Forum thread recommendation for massive open online courses," in *Proceedings of 7th International Conference on Educational Data Mining*, 2014.
- [16] D. Yang, D. Adamson, and C. P. Rosé, "Question recommendation with constraints for massive open online courses," in *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*, 2014, pp. 49–56.
- [17] M. Jenders, R. Krestel, and F. Naumann, "Which Answer is Best? Predicting Accepted Answers in MOOC Forums."
- [18] A. Agrawal, J. Venkatraman, S. Leonard, and A. Paepcke, "YouEDU: Addressing Confusion in MOOC Discussion Forums by Recommending Instructional Video Clips," 2015.
- [19] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The adaptive web*, Springer, 2007, p. 291.
- [20] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, p. 5, 2004.
- [21] G. Shani and A. Gunawardana, "Evaluating recommendation systems," in *Recommender systems handbook*, Springer, 2011, pp. 257–297.
- [22] G. Schröder, M. Thiele, and W. Lehner, "Setting Goals and Choosing Metrics for Recommender System Evaluations," in *UCERSTI2 Workshop at the 5th ACM Conference on Recommender Systems, Chicago, USA*, 2011, vol. 23.

