

Diseño de sistemas recomendadores en el ámbito de las bibliotecas universitarias

Sergio A. Cabrera, Cristian D. Pacífico y Juan C. L. Teze

Área de Agentes y Sistemas Inteligentes, Facultad de Ciencias de la Administración,
Universidad Nacional de Entre Ríos,
Monseñor Tavella 1424, (E3202KAC) Concordia
{sergio.cabrera, cristian.pacifico, carlos.teze}@uner.edu.ar

Abstract. Esta línea de I+D se ubica dentro del grupo de investigación Agentes y Sistemas Inteligentes de la Facultad de Ciencias de la Administración dependiente de la Universidad Nacional de Entre Ríos (FCAD/UNER).

Persigue como propósito, diseñar un modelo de sistema que utilice técnicas de recomendación y de aprendizaje automático, para realizar sugerencias de fuentes de información, en el ámbito de las bibliotecas universitarias.

En esta etapa de la investigación, se analizarán las técnicas de recomendación, en especial las basadas en contenido, en el ámbito de las bibliotecas universitarias.

Keywords: Sistemas Recomendadores, Bibliotecas Universitarias, Machine Learning.

1. Bibliotecas y Bibliotecas Universitarias

La biblioteca es una organización que tiene como misión fundamental la de otorgar acceso a la información a diferentes usuarios que hacen uso de su acervo documental. Las bibliotecas universitarias se caracterizan por brindar acceso y difusión de fuentes de información, colaborando en los procesos de creación de conocimiento, sirviendo de apoyo a la educación y a la investigación, tanto a la comunidad académica en la que se encuentra inmersa, como así también a la sociedad en general.

La aplicación de la tecnología computacional a las etapas de la actividad científico informativa, ha favorecido la utilización de los Sistemas Integrados de Gestión Bibliotecaria (SIGB).

Se puede definir un SIGB, como un conjunto de programas informáticos diseñados para atender los procesos de las bibliotecas, brindando soluciones a los requerimientos de las mismas, ayudando en los procesos técnicos y en los servicios prestados, favoreciendo la intercomunicación entre bibliotecas. Dentro de estos sistemas se destaca KOHA, por ser el primer SIGB de software libre, permitiendo trabajar con normas de catalogación internacionales, utilizando el formato MARC y el protocolo Z39.50 para la recuperación de información [1].

La Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA)[2], ha enunciado una serie de políticas y recomendaciones con el objetivo de redefinir el trabajo de los bibliotecarios y el rol que las bibliotecas desempeñan en una sociedad que cada día más se amalgama con la Inteligencia Artificial (IA) y en especial con Machine Learning (ML). Las técnicas de IA y ML podrían integrarse con el trabajo

diario del bibliotecario, permitiendo potenciar los servicios y las funciones de las bibliotecas.

2. Sistemas recomendadores

La mayor parte de las actividades que desarrollamos a diario necesitan nutrirse de información para tomar decisiones. El gran volumen de datos e información que se generan diariamente, hacen que esta tarea se torne compleja.

Según Maes, Jannach, Ricci y Melville, como se citó en Briguez [3], los sistemas de recomendación (SSRR) son mecanismos de apoyo que ayudan a los usuarios en su proceso de toma de decisiones al interactuar con volúmenes de información grandes o complejos. La mayoría de los SSRR están destinados a ayudar a los usuarios a hacer frente al problema de la sobrecarga de información, facilitando el acceso a los elementos pertinentes.

Ricci, como se citó en Charnelli [4], establece que la tarea de un Sistema Recomendador (SR) es seleccionar automáticamente los artículos más apropiados para cada usuario de acuerdo a sus intereses y preferencias personales. Generalmente, un SR se centra en un tipo específico de elemento a recomendar denominado “ítem”, como por ejemplo un repositorio de materiales educativos, que genera recomendaciones personalizadas para proporcionar sugerencias útiles y efectivas para ese tipo específico de ítem. Se debe tener en cuenta que dependiendo del enfoque de recomendación, según el contexto y la necesidad del usuario, los SSRR generan sugerencias utilizando diversos tipos de conocimiento y datos sobre los usuarios, los ítems disponibles y las transacciones anteriores almacenadas en bases de datos personalizadas.

Según Charnelli, los SSRR se pueden clasificar en cinco categorías dependiendo de la técnica de recomendación: basados en el contenido, filtrado colaborativo, basados en conocimiento, basados en comunidad y por último los híbridos.

Estas técnicas de recomendación pueden mejorarse con las desarrolladas para ML. El ML es un subcampo de las ciencias de la computación y una rama de la IA, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. El aprendizaje mejora, cuando un sistema logra optimizar su desempeño con la experiencia y el uso de datos, que se convierten en información necesaria para la toma de decisiones.

Los avances en este sentido agregarían nuevas dimensiones y enfoques a los procesos de gestación del conocimiento que se producen en las bibliotecas, en particular, los referidos a la organización, el almacenamiento y la integración de conocimiento. En este sentido se han propuesto soluciones que integran modelos producidos por algoritmos de ML a la construcción de SSRR. De esta forma, se lograron mejorar determinados componentes, haciendo más precisa la recomendación.

Entre algunos desarrollos actuales, es posible citar a Ex Libris Alma [5], como una plataforma unificada de servicios de biblioteca. DARA es el motor inteligente que da soporte a las decisiones de Alma, dotando de IA a las bibliotecas; utiliza algoritmos y análisis inteligentes para adaptar sus recomendaciones a las características de cada una. Además cabe mencionar a Hamlet, que es un sistema que utiliza ML promoviendo interfaces exploratorias experimentales para la colección de Tesis del Instituto

Tecnológico de Massachusetts (MIT). Está en desarrollo y algunas interfaces no se han finalizado todavía. [6][7].

3. Solución propuesta

En primera instancia se pretende caracterizar las técnicas de recomendación, realizando una síntesis de las mismas que asocie a cada una con sus dominios de aplicación más adecuados.

Al particularizar la función de recomendación, se espera obtener un esquema de requisitos y de características deseables de un SSRR de fuentes de información. Luego se realizará una síntesis de las técnicas de ML y su forma de aplicación en el contexto de recomendación de fuentes de información.

3.1 Particularización de la función de recomendación

De la especificación de los SIGB, se identificarán las funcionalidades y los componentes esenciales, resultando de especial interés el estudio de las funciones para el servicio de referencia y el de diseminación selectiva de la información.

En esta etapa de la investigación, se están analizando las técnicas de recomendación en el ámbito de las bibliotecas universitarias, para lo cual se considerarán ciertas particularidades correspondientes a los usuarios y a las fuentes de información.

Se trabajará sobre los usuarios en la categoría de estudiantes. Para confeccionar el perfil de los mismos, serán relevantes los datos provistos por el sistema de gestión académica, SIU-GUARANÍ que se detallan a continuación: edad, carrera, año, materias que cursa, materias aprobadas. Además, será necesario contar con los datos provistos por el historial de préstamos suministrados por SIGB KOHA.

En cuanto a las fuentes de información, se considerarán libros de los cuales serán necesarios los siguientes datos: autor, título, año de edición y cantidad de préstamos realizados durante el último año. Estos datos serán provistos por el SIGB KOHA.

3.1.1 Sistema de recomendación basado en contenido

En estos sistemas, las recomendaciones se basan en el principio de contenido similar. Si un usuario está leyendo un libro, el sistema recomendará otros libros de contenido similar. Se utilizan varios atributos fundamentales para calcular la similitud al verificar contenido semejante.

Esta técnica toma en cuenta algunos datos del historial del usuario para intentar predecir qué busca y cuáles son las sugerencias similares que puede mostrar.

Si bien estos recomendadores presentan algunos inconvenientes, tales como: la sobreespecialización, el análisis de contenido limitado o, en menor medida, el problema de un nuevo usuario (arranque en frío), los mismos deben ser estudiados en el ámbito de las bibliotecas universitarias, dado que se pueden descubrir opciones que se ajustan a las características de los contenidos que se han seleccionado con anterioridad,

permitiendo elegir elementos similares nuevos. Además, los SSRR basados en contenido, son uno de los que presenta mayor presencia en la actualidad.

Para explicar cómo funcionan, se presentará el siguiente caso:

Si un estudiante que se encuentra cursando la asignatura Estructura de Datos, en la carrera de Licenciatura en Sistemas, necesitara solicitar en préstamo un libro de estructura de datos en Java, del autor Mark Allen Weiss; luego de consultar el catálogo de la biblioteca de la Facultad de Ciencias de la Administración, las posibilidades serían:

- Weiss, Mark Allen. Estructura de datos y algoritmos en Java. Ed. 2000. Cantidad de páginas: 800. Cantidad de préstamos último año: 5.
- Weiss, Mark Allen. Estructura de datos y algoritmos en Java y C++. Ed. 2013. Cantidad de páginas: 1000. Cantidad de préstamos último año: 100.
- Weiss, Mark Allen. Estructura de datos en Java. Ed. 2013. Cantidad de páginas: 995. Cantidad de préstamos último año: 102.

Para que el SR pueda sugerir un libro, deberá comprobar la similitud entre los ejemplares. Para determinarla se utilizará el método de distancia euclidiana, en función del año de edición, la cantidad de páginas y la cantidad de préstamos en el último año. Distancia euclidiana entre el libro a y el c:

$$\sqrt{(2000 - 2013)^2 + (800 - 995)^2 + (5 - 102)^2} = \sqrt{(-13)^2 + (-195)^2 + (-97)^2} = \sqrt{47603} = 218,18$$

Distancia euclidiana entre el libro b y el c:

$$\sqrt{(2013 - 2013)^2 + (1000 - 995)^2 + (100 - 102)^2} = \sqrt{(0)^2 + (-5)^2 + (-2)^2} = \sqrt{29} = 5,38$$

Se puede observar que la distancia euclidiana es menor entre los ejemplares b y c, que entre los a y c, lo que significa que entre los primeros hay más similitudes que entre los segundos, por lo tanto, el SR, sugeriría los libros b y c.

Para poder explicar el concepto a través de este ejemplo, solo se tomaron algunas características básicas de los libros, pero no existen restricciones para realizar el cálculo de la distancia sobre las características del producto. El principio básico sigue siendo el mismo, si la distancia entre ambos es 0, es probable que tengan un contenido muy similar. Para lograr recomendaciones que se aproximen a las necesidades de los usuarios, lo ideal es combinar diferentes técnicas.

Mathew, Kuriakose y Hegde [8], proponen una estrategia de mejora para los sistemas de recomendación de libros mediante la combinación en cascada de dos métodos: el filtrado basado en contenido y el filtrado colaborativo, logran así que el sistema presente los resultados obtenidos como recomendaciones personalizadas para el usuario. Por su parte, Xin, Haihong, Junde, Meina y Junjie [9], combinan el filtrado colaborativo de vecinos cercanos y el modelo de factor latente. Posteriormente, se exploraron dos modelos para fusionar estas técnicas: la regresión lineal y las redes neuronales. Los resultados indicaron que obtener mejores resultados implicaba combinar distintas técnicas en lugar de implementarlas por separado. Además, destacaron que se lograron predicciones más precisas al emplear redes neuronales en el proceso.

Por otro lado, Tsuji, Kuroo, Sato, Ikeuchi, Ikeuchi, Yoshikane y Isumura [10], llevaron a cabo un estudio comparativo de tres técnicas de recomendación de libros: el filtrado colaborativo, la minería por reglas de asociación y el algoritmo de Amazon. Su objetivo era determinar cuál de estos métodos era el más efectivo y qué atributos debían ser considerados en el proceso de recomendación. Concluyeron que el algoritmo de Amazon y la minería por reglas de asociación son opciones más efectivas para la recomendación de libros. Se resalta la importancia de considerar los aspectos de privacidad y el costo de los sistemas de recomendación basados en filtrado colaborativo.

De la caracterización de las técnicas de recomendación, se espera lograr una síntesis de las técnicas de recomendación que asocie a cada una con las sugerencias a realizar en el ámbito de las bibliotecas universitarias.

4. Líneas futuras

Luego de realizar un análisis comparativo de las soluciones existentes, realizando pruebas experimentales, se estará en condiciones de diseñar y especificar una arquitectura general para SSRR, proponiendo una guía metodológica para su implementación.

Cabe destacar que la experiencia de un grupo de investigación consolidado en el área de investigación Agentes y Sistemas Inteligentes, habilitará la formación de recursos humanos en la FCAD/UNER, debido a que en paralelo al desarrollo de la investigación, se llevará a cabo la redacción de una tesis de maestría y una tesina de grado.

Referencias

1. KOHA. Consultado el 10 de marzo de 2023. <https://koha-community.org/>
2. IFLA (2020). IFLA Statement on Libraries and Artificial Intelligence. <https://repository.ifla.org/handle/123456789/1646>
3. Briguez, C. E. Integración de técnicas cualitativas y cuantitativas en los sistemas de recomendación. [Tesis de Doctor en Ciencias de la Computación, Universidad Nacional del Sur]. Repositorio institucional <https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/4736/BRIGUEZ%2C%20Cristian%20Emanuel%20-%20Tesis%20Doctoral.pdf?sequence=5&isAllowed=y>
4. Charnelli, M. E. (2019). Sistemas recomendadores aplicados en educación. [Tesis Especialización en Tecnología Informática Aplicada en Educación, Universidad Nacional de La Plata]. Repositorio institucional <http://sedici.unlp.edu.ar/handle/10915/67063>
5. ExLibris. Consultado el 10 de febrero de 2023. <https://exlibrisgroup.com/>
6. American Library Association. (2019). Artificial Intelligence and Machine Learning in Libraries. Library Technology Reports, 55(1). <https://doi.org/10.5860/ltr.55n1>
7. Yelton, A. (s.f.). Hamlet. Consultado el 18 de febrero de 2023. <https://hamlet.andromedayelton.com/>
8. Mathew, P., Kuriakose, B. y Hegde, V. (2016). Book recommendation system through content based and collaborative filtering method, International Conference On Data Mining And Advanced Computing (SAPIENCE).

9. Xin, L., Haihong, E., Junde, S, Meina, S. y Junjie, J. (2013). Collaborative book recommendation based on readers borrowing records. International Conference On Advanced Cloud And Big Data.
10. Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F. y Itsumura, H. (2012). Use of library loanrecords for book recommendation, IIAI International Conference On Advanced Applied Informatics.