

Análisis, diseño y desarrollo de un sistema de recomendación basado en datos restauranteros de TripAdvisor y Foursquare

Saúl Pérez, Mary Carmen Cuecuecha, José Federico Ramírez,
José Crispín Hernández

Instituto Tecnológico de Apizaco, División de Estudios de Posgrado e Investigación,
Apizaco, Tlaxcala,
México
{saul.perez.tirzo, MaryCarmenCuemu, federico.ramirez}@gmail.com,
josechh@yahoo.com

Resumen. Debido a la inmensa cantidad de información disponible en Internet, provoca que los usuarios se sientan abrumados con tanta información, haciendo difícil el proceso de búsqueda de productos y/o servicios que se adecuen a los gustos y necesidades de cada usuario. Por esta razón el desarrollo de herramientas inteligentes se ha vuelto indispensable como lo son los Sistemas de Recomendación, donde su objetivo primordial es ayudar a los usuarios a encontrar información de productos y/o servicios de mejor manera filtrando toda la información disponible logrando así un mejor uso de ella. En el presente trabajo de investigación se diseña y desarrolla un Algoritmo de Recomendación Híbrido para crear una lista de ítems (restaurantes) recomendables a los usuarios (consumidores), fusionando los algoritmos: Filtro Colaborativo y Basado en Contenido, utilizando un Clasificador Bayesiano con técnicas de Procesamiento de Lenguaje Natural. Además se mejora la experiencia del usuario aplicando la ubicación GPS del usuario como un filtro a las recomendaciones. Para medir el rendimiento del sistema propuesto se experimentó con un conjunto de datos extraídos de los Sitios Web Foursquare y TripAdvisor.

Palabras clave: Sistema de recomendación híbrido, filtro colaborativo, filtro basado en contenido, procesamiento de lenguaje natural, clasificador bayesiano.

Analysis, Design and Development of a Recommendation System Based on Tripadvisor and Foursquare Restaurant Data

Abstract. Due to the immense amount of information available on the Internet, it causes users to feel overwhelmed with so much information, making it difficult to search for products and/or services that suit the tastes and needs of each user. For this reason the development of intelligent

tools has become indispensable as are the Recommendation Systems, where its main objective is to help users find information of products and/or services in a better way filtering all the available information thus achieving a better use of it. In the present research work a Hybrid Recommendation Algorithm is designed and developed to create a list of recommended items (restaurants) for users (consumers), merging the algorithms: Collaborative and Content Based Filter, using a Bayesian Classifier with techniques of Natural Language Processing. Besides, the user experience is improved by applying your GPS location as a filter to the recommendations. To measure the performance of the proposed system, we experimented with a set of data extracted from the Foursquare and TripAdvisor Websites.

Keywords: Hybrid recommendation system, collaborative filtering, content-based filtering, natural language processing, Bayesian classifier.

1. Introducción

La industria restaurantera se encuentra en una fase completamente nueva, donde es necesario el uso de Tecnologías de Información y Comunicaciones (TICs) de vanguardia que el mundo globalizado actual demanda para ser competitivos dentro del mercado, representando uno de los principales sectores en pro de la economía [14], además de promover el turismo de la región, pretende que sus consumidores satisfagan una de sus necesidades básicas, pero de las más importantes, la alimentación.

En esta era de la web hay una sobrecarga de información excediendo la capacidad de una persona para poder procesarla en términos de lo que un usuario necesita encontrar; particularmente la búsqueda de restaurantes que tengan características o platillos peculiares para un determinado consumidor resulta una tarea tediosa y a menudo difícil, debido a que los consumidores buscan lugares apropiados a sus gustos personales, pero gracias a la gran cantidad de información que se puede encontrar en internet los resultados obtenidos no siempre son los mejores, lo que genera una satisfacción parcial, siendo una de las razones principales por las que los Sistemas de Recomendación (SR) juegan un rol importante en nuestra vida cotidiana [10], hasta cierto punto, el problema está siendo resuelto por los motores de búsqueda, pero no proporcionan la personalización de los datos. La personalización ha sido reconocida como un factor crítico para las industrias restauranteras exitosas y la utilización de sistemas de recomendación es el mejor enfoque para tratar con el problema de la personalización [21].

En un inicio los Sistemas de Recomendación emergen como una área de investigación individual cuando algunos investigadores iniciaron a trabajar en diferentes problemas de recomendación [23].

Los Sistemas de Recomendación (SRs) tienen una relación con los Sistemas de Búsqueda o Recuperación de Información, dado que ambos están diseñados

para que a partir de un conjunto de datos se obtenga información relevante para el usuario [1]. De acuerdo con Ricci [26] dice que un Sistema de Recomendación (SR) puede brindar información valiosa para asistir en el proceso de toma de decisiones del consumidor con el objetivo de proveer una recomendación de restaurantes con exactitud y de manera precisa [21], ya que las recomendaciones que se producen en un Sistema de Recomendación, se reducen a ayudar a los seres humanos a satisfacer sus gustos personales y descubrir nuevos elementos, con menos esfuerzo, que si realizaran la actividad de manera manual. Para lograr esto, computacionalmente, los SRs recolectan las preferencias de los usuarios, de forma explícita, por ejemplo: solicitando a los usuarios valoraciones sobre los elementos, o infiriéndolas a partir de las acciones de los usuarios. En la actualidad, los SRs han demostrado ser un medio valioso para lidiar con el problema del exceso de información [21]

Este trabajo presenta las bases para el desarrollo de un Sistema de Recomendación Híbrido de acuerdo a la clasificación propuesta por Burke [4], además de implementar el uso de técnicas de Inteligencia Artificial (IA) que permitan sugerir recomendaciones efectivas de restaurantes, orientados a mejorar la toma de decisiones de una manera sencilla y en un tiempo considerable, de acuerdo a los gustos de cada consumidor. La motivación de este trabajo se concentra en establecer la investigación, diseño y desarrollo de un Sistema de Recomendación para la industria restaurantera. La estructura del trabajo se conforma de la siguiente forma: sección 2 Trabajos relacionados, sección 3 Método propuesto, sección 4 Evaluación experimental y sección 5 Conclusiones y Trabajos a futuro.

2. Trabajos relacionados

Actualmente ha crecido el área de los SRs debido a la evolución constante de plataformas digitales que requieren de tecnologías inteligentes para generar conocimiento a partir de las grandes cantidades de datos que se generan. Jain, Grover, Thakur y Choudhary [16] ejemplificaron a Youtube.com, LinkedIn.com y Amazon.com como los principales sitios de internet que operan con un sistema de recomendación. Es importante remarcar que los sistemas de recomendación se han convertido en una herramienta de vital importancia para los usuarios, ya que engloban una infinidad de soluciones que tienen como objetivo ayudar en la selección de productos y/o servicios con el fin de facilitar la decisión de compra y consumo, y reducir el tiempo empleado mejorando la experiencia del usuario.

Hoy en día, los sistemas de recomendación juegan un papel importante en todos los campos, especialmente en el restaurante, centros de comida o el turismo [29]. Da Costa, Suyoto y Joko [5] propusieron un sistema de recomendación híbrido (Filtro Colaborativo y Basado en Contenido), con la implementación del algoritmo de vecinos cercanos (k-NN) y servicios de ubicación para brindar recomendaciones a los turistas y a su vez guiarlos a las mejores atracciones turísticas en Timor Oriental.

Las atracciones turísticas se basan en varias categorías: Comida/Gastronomía, Cultural, Historia y Religioso, por mencionar algunas. Los resultados obtenidos

revelan un interesante aporte a la investigación ayudando de manera eficiente a los turistas a planificar su destino en Timor Oriental.

Por otro parte Ananta, Suyoto, Joko [29] presentaron un sistema de recomendación utilizando el enfoque colaborativo con distancia euclidiana como el orden de la similitud de la calificación de los usuarios, para recomendar restaurantes a los turistas que llegan a Buleleng Regency (Indonesia), también se utiliza la tecnología de servicio basado en la ubicación, que utiliza el posicionamiento global (GPS) para que los usuarios puedan conocer su posición, definir y buscar ubicaciones específicas, ya sea lejos o cerca, con el propósito de encontrar el centro de comida adecuado en Buleleng Regency. Es importante mencionar que este sistema se ejecuta en dispositivos móviles Android. Por último los autores mencionan que para una mejor recomendación es necesario implementar otros enfoques que tomen en consideración los intereses del usuario. Martínez, Rodríguez y Espinilla [20] desarrollaron un sistema de recomendación híbrido para restaurantes de la región de Jaén (España), empleando el filtro colaborativo y basado en conocimiento, evitando de esta forma el problema de arranque en frío. Además introdujeron un módulo para sus usuarios que consiste en la información geográfica que proporciona Google Maps sobre los restaurantes recomendados, siendo capaz de proporcionar recomendaciones en cualquier situación requerida por los usuarios.

Otro trabajo de gran interés fue el propuesto por Habib, Rakib y Abul [13], quienes remarcaron que las redes sociales basadas en la ubicación, representan una plataforma para comprender las preferencias de los usuarios, abriendo un campo prometedor dentro de la investigación. Estudiaron la amalgamación (fusión) de registros de GPS en tiempo real, además de los datos históricos de check-in en el desarrollo de un sistema de recomendación para restaurantes, su método de recomendación toma en consideración la hora del día, la ubicación actual del usuario, sus preferencias y el historial de registro del usuario, el cual se analiza individualmente para descubrir tendencias de visitas de los usuarios, tendencias de preferencia de los alimentos y la popularidad general de los restaurantes, donde los resultados experimentales confirmaron la efectividad del método propuesto. Por otro lado Ketmaneechairat, Kongketwanich y Najit [17] diseñaron y desarrollaron un sistema de recomendación para la cocina de comida tailandesa en teléfonos inteligentes. El desarrollo de la aplicación fue bajo la plataforma Android.

El objetivo perseguido con este sistema fue el brindar a los usuarios interesados información detallada y certera acerca de la cocina tailandesa. Algo interesante de este trabajo es que los resultados se muestran en dos idiomas: tailandés e inglés, además de detalles adicionales de los alimentos como el nombre del platillo, su imagen, los ingredientes que lo componen y su método de cocción. La aplicación consiste de dos funciones de búsqueda: buscar usando la categoría de la comida y buscar usando una palabra clave. Los resultados de las recomendaciones generadas por el sistema mostraron la relevancia del trabajo, ya que hubo una versión disponible en PlayStore para usuarios interesados en probar su funcionamiento.

Los SRs se clasifican en diferentes categorías de acuerdo al tipo de información que se utiliza para recomendar productos y/o servicios a los usuarios [18]. En la literatura se puede encontrar una diversidad de técnicas para implementar los SRs por ejemplo en la Fig. 1 se muestran las técnicas más usadas [11].

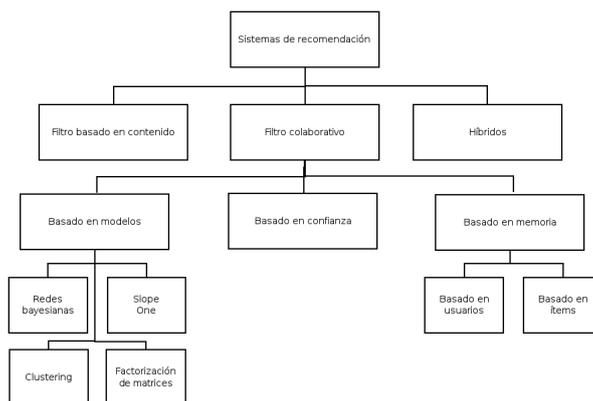


Fig. 1. Clasificación de los SRs.

Los principales enfoques que se emplean es el Filtro Colaborativo y el Filtro Basado en Contenido, pero ambos tienen algunas limitaciones y problemas. Mansur y Patel [18] en su trabajo de investigación dan un panorama general de los Sistemas de Recomendación, donde describieron los principales enfoques, problemas y limitaciones existentes. Así que en la mayoría de las aplicaciones que ocupan un SR se utiliza un enfoque híbrido, que combina diferentes técnicas para mejorar el funcionamiento del sistema, donde su idea principal es generar recomendaciones con una mejor exactitud y eficiencia; en cambio cuando se aplica un solo algoritmo es posible obtener resultados inexactos y poco confiables.

Como se mencionó anteriormente los sistemas de recomendación están presentes en muchos sectores productivos, además del restaurantero. En la literatura se encuentran diversas aplicaciones de un SR Híbrido por ejemplo, Pal, Parhi y Aggarwal [22], desarrollaron un algoritmo híbrido que donde fusionaron las ventajas del Filtro Colaborativo y el Filtro Basado en Contenido para generar recomendaciones de películas, el método presentado se enfoca en encontrar la correlación entre dos características usando la intersección de conjuntos en el filtro basado en contenido, y encontrar la similitud entre dos elementos (películas) y predecir las recomendaciones. El algoritmo fue probado y comparado con el filtro colaborativo puro (PCF) y SVD (Singular value Descomposition). Los valores MAE que se generaron después de la evaluación proporcionaron comparaciones exitosas, siendo el sistema híbrido el que produjo mejores valores MAE, mejorando la dispersión del conjunto de datos entre 1 %-2 %.

El análisis que presentaron Bogers y Van den Bosch [3] en su estudio incorporaron dos enfoques diferentes: filtro basado en contenido y filtro colaborativo

para recomendar artículos en los sitios web de marcado social de libros, incorporando las etiquetas y metadatos (títulos, descripciones, etc.) que se generan en los sitios web de marcadores sociales, los cuales permiten a los usuarios almacenar, organizar y buscar marcadores de páginas web. El objetivo primordial de esta investigación fue el predecir qué marcadores ocultos pueden gustarle al usuario en base a su perfil. Generaron recomendaciones directamente basadas en la divergencia de Kullback-Leibler de los modelos de lenguaje de metadatos, exploraron el uso de esos metadatos para calcular la similitud entre usuarios y elementos. Los experimentos realizados fueron con tres conjuntos de datos de dos dominios diferentes: Delicious, CiteULike y BibSonomy.

Por otro lado, Renjith y Anjali [24] mostraron un SR usando un algoritmo híbrido para generar recomendaciones personalizadas a los usuarios durante su viaje, compras y otras rutinas que realice transformándolo en un compañero esencial. La aplicación es un recomendador de viajes híbrido que reúne las características de las técnicas de filtro basadas en contenido, colaborativo y demográfico, para extraer el patrón de comportamiento del usuario, permitiendo hacer recomendaciones adecuadas. El patrón de comportamiento aprende de las huellas de viaje del usuario, por lo tanto, no se necesita la intervención del usuario para generar la predicción y las recomendaciones personalizadas. Cabe mencionar que este trabajo de investigación resultó muy interesante, además los autores mencionaron las posibles mejoras que podrían ser implementadas en el sistema para afinar las recomendaciones como: medios de transporte para llegar a la atracción recomendada, que además de económico sea confiable. Instalaciones de alojamiento con ofertas interesantes y parámetros climáticos externos.

Además Ashraf, Rabbani, Martinez y Muhammad [2] propusieron un sistema de recomendación para los servicios en la nube de manera que el usuario pueda contratar el servicio que cumpla con sus propios requerimientos, lo que hace que sea fácil para un usuario decidir. El objetivo perseguido fue probar los servicios de acuerdo con las necesidades del usuario y algunos parámetros establecidos como: la seguridad, la privacidad, el almacenamiento de datos, etc. y así evaluar los servicios. Además de cumplir con los requisitos máximos del usuario, por lo que el tiempo de navegación disminuye y el buscador de servicio puede aprovechar el mejor de todos los servicios en la nube disponibles.

Los autores creen conveniente que las opiniones de los usuarios también son importantes de acuerdo a sus fechas para que se defina el concepto y el funcionamiento de un servicio en un momento determinado. García, Tello, De la Rosa y Sánchez [10], presentaron un Sistema de Recomendación para Música que aplica técnicas de aprendizaje semi-supervisado, que es capaz de aprender y adaptarse a los gustos de sus usuarios sin la necesidad de tener información previa del perfil de usuario y determinar cuándo una canción es probablemente atractiva para un usuario en particular. Los algoritmos de aprendizaje utilizados fueron Naive Bayes y J48. Para sus pruebas experimentales utilizaron un subconjunto de datos extraídos de la base de datos musical Gracenote, donde sus resultados obtenidos mostraron que con un conjunto reducido de características es posible construir de forma efectiva un modelo de recomendación. Para evaluar el método

que propusieron utilizaron las medidas tradicionales de evaluación: precisión, recall y medida F. Concluyeron que los árboles de decisión (J48) obtuvieron mejores resultados de clasificación y al mismo tiempo decir que el atributo de género es el más importantes en el proceso de clasificación.

Guzmán y Torres [12] propusieron las bases para un sistema de recomendación móvil que se basa en técnicas de Inteligencia Artificial, permitiendo sugerir a un usuario rutinas de ejercicio, orientadas a fortalecer la calidad de vida del usuario basándose en su perfil antropomórfico y patológico. Además la arquitectura de cómputo en la nube, permite el acceso libre a la aplicación y visualización de videos en el dispositivo móvil. El sistema de recomendación trabaja en función de la interpretación de la información suministrada por el usuario, ubicando al individuo dentro de un grupo poblacional identificado por patologías comunes y así asignarle una rutina adecuada. Cada rutina se generó con la información que el usuario proporciona, construyendo de esta manera un sistema de reglas que permita razonar y recomendar rutinas de ejercicios idóneas, siendo un aporte interesante para la los campos de la investigación y la salud.

Y finalmente Otero, Gonzales y Franco [21] esbozaron un sistema de recomendación colaborativo usando un SIG (Sistema de Información Geográfica) para sugerir productos turísticos en Cuba, el cual consta de una interfaz gráfica de usuario y un motor de recomendación, donde los datos son procesados para generar las recomendaciones. Para pruebas experimentales utilizaron una base de datos con 4000 usuarios y 1000 productos turísticos. Analizando la información que brindó la herramienta se emplearon indicadores para la calidad de las recomendaciones obtenidas, como lo son la probabilidad de recomendaciones relevantes (precisión) y la probabilidad de que los ítems relevantes sean recomendados (recall). Finalmente los autores concluyen que las pruebas realizadas demostraron que para el contexto turístico descrito se obtuvieron resultados positivos en la exactitud y el uso de las predicciones de valoración, calculadas mientras se incide positivamente en la cobertura del sistema implementado, lo que permite afirmar que la aplicación de la técnica logró obtener de manera eficaz el contenido que debe ser mostrado prominentemente a los usuarios de las organizaciones de turismo en Cuba.

Con lo descrito anteriormente se concluye que el campo de investigación de los sistemas de recomendación es amplio y diverso, ya que no solo se basan en los servicios que existen en internet sino su popularidad de esta clase de sistemas ha generado que su implementación sea en diferentes áreas de la vida real lo que conlleva a desafíos interesantes para estudios futuros.

3. Método propuesto

El SR propuesto utiliza un enfoque híbrido que toma como componentes principales: los usuarios que desean obtener sugerencias de establecimientos para comer y los ítems que son representados por los restaurantes.

En [15] se describen siete técnicas para implementar un SR Híbrido: *Weighted*, *Switching*, *Mixed*, *Feature*, *Feature augmentation*, *Cascade* y *Meta level*;

en este método se utiliza la técnica Switching, debido a que el sistema utiliza la situación actual del usuario. Por ejemplo, si el usuario no cumple con las condiciones para emplear el Filtro Colaborativo, entonces se aplica el Filtro Basado en Contenido. El objetivo de este Sistema de Recomendación Híbrido es cubrir las necesidades que tengan los diferentes usuarios.

3.1. Colección de datos

Debido a la falta de los datos idóneos a los propósitos planteados, fue necesaria la construcción de un conjunto de datos que permitirá evaluar la pertinencia del método propuesto. Para esto se obtuvieron datos del Sitio Web Foursquare, una compañía de tecnología que utiliza la ubicación inteligente para construir una experiencia significativa para el consumidor [7]. Otro Sitio Web que también fue utilizado es TripAdvisor, un sitio de viajes que permite a los viajeros aprovechar al máximo el potencial de cada viaje proporcionando reseñas de contenido relacionado con los viajes [27].

El sitio Foursquare proporciona una API¹ (Interfaz de Programación de Aplicaciones, abreviada como *API* del inglés: *Application Programming Interface*) que permite consumir un servicio web para extraer datos sobre lugares (restaurantes, cafeterías, bares, etcétera)². Para este trabajo se obtuvieron los lugares de las categorías *food*, *drinks* y *coffee* [8] tomando de muestra los que se encuentran en el Estado de Tlaxcala. Como resultado se concentraron 685 lugares seleccionando solo algunos datos, además se obtuvieron 4764 comentarios relacionados con los lugares y 65274 calificaciones registradas por 53051 usuarios. En la Tabla 1 se describen los datos que fueron seleccionados para concentrar la información.

Tabla 1. Datos seleccionados de la API de Foursquare sobre los lugares.

Dato	Descripción
id	Un identificador único en formato de cadena.
name	Nombre establecido.
lat	Valor de la latitud para indicar la ubicación.
lon	Valor de la longitud para indicar la ubicación.
type	Categoría.
text	Texto del comentario.
user_id	Un identificador único para el usuario.
rating	Calificación otorgada al lugar por el usuario.

¹ La API es un conjunto procedimientos que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

² Para mayor información puede visitar este sitio web <https://developer.foursquare.com>

Debido a la escasez de los comentarios obtenidos en Foursquare se opta por utilizar la técnica Web Scraping que sirve para extraer información de páginas web de forma automatizada [19]. Para implementar la técnica, se utilizó la herramienta scrapy³ que funciona con el lenguaje de programación Python. Dicha técnica se aplicó al sitio TripAdvisor para conseguir un total de 38107 comentarios de los cuales se etiquetaron de acuerdo a la calificación otorgada por el usuario quedando como resultado 14876 comentarios y con estos se categorizaron en dos clases: 7438 positivos y 7438 negativos.

Los datos se pueden descargar desde este enlace https://drive.google.com/drive/folders/1XLyn-of_Y9zch9jmf81nCbKvFAarA19.

3.2. Descripción del sistema de recomendación híbrido

La idea principal del sistema de recomendación híbrido es implementar las técnicas más populares que existen, el Filtro Colaborativo y Basado en Contenido. Además se integra la ubicación del usuario con el sistema de posicionamiento global más conocido por sus siglas en inglés, GPS (siglas de Global Positioning System)[9] para generar sugerencias de relevancia, porque no resultan relevantes aquellos lugares que se encuentren lejos del usuario.

En la Fig. 2 se observan los componentes principales del sistema de recomendación híbrido.

La propuesta inicia con el usuario, cuando registra información personal como es la calificación que le otorga a cada restaurante y los principales gustos que tiene, esto se almacena en una base de datos para su uso posterior. Después se obtiene el historial del usuario con los datos registrados previamente, evidentemente que a través del tiempo todos estos datos empezarán a generar recomendaciones significativas. Entonces en este punto es cuando empieza el trabajo de los diferentes algoritmos que integran al SR Híbrido. Como primer paso se valida la información que ha generado el usuario; si el usuario tiene calificaciones registradas, entonces se obtienen todas las que fueron otorgadas por él a los restaurantes y se utilizan en el Algoritmo del Filtro Colaborativo.

En el trabajo propuesto por Yang, Wu, Zheng, Wang y Lei [30] mostraron los principales Algoritmos de Filtro Colaborativo: Basado en el Usuario y Basado en el Ítem. El Sistema de Recomendación propuesto, utiliza el Algoritmo Basado en el Ítem debido a que el número de usuarios es más largo que el número de ítems y los ítems no tienen cambios frecuentemente. La Fig. 3a se presenta el proceso que emplea el algoritmo del Filtro Colaborativo para generar una lista previa de recomendaciones.

El primer paso es calcular la similitud que existe entre los ítems mediante la matriz de calificaciones que se crea con la información de la base de datos. Como ejemplo de la matriz de calificaciones se presenta en la Tabla 2.

Para el cálculo de la similitud se pueden utilizar las siguientes medidas: puro coseno, coseno ajustado o coeficiente de correlación Pearson [30]. En este trabajo

³ Es una herramienta open source para extraer datos des sitios web de forma rápida, simple y completa. <https://scrapy.org>

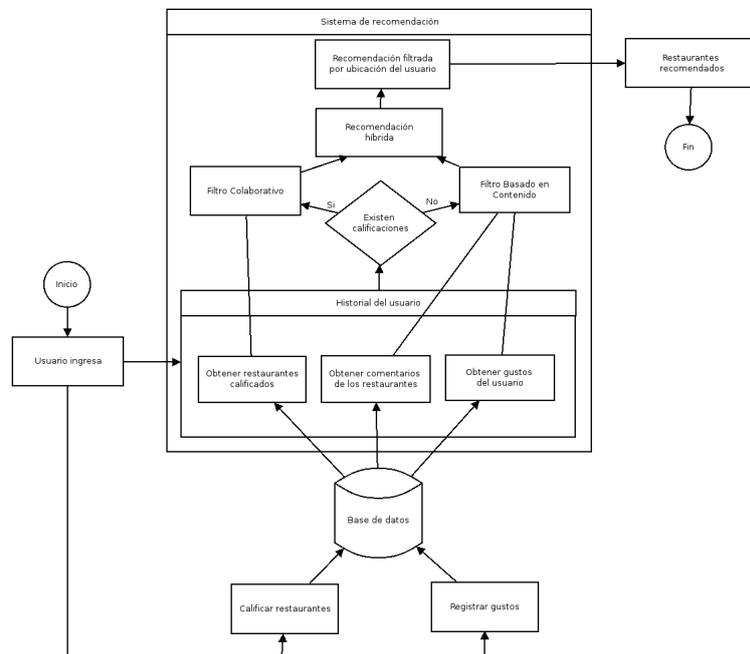


Fig. 2. Diagrama de bloques del sistema de recomendación híbrido.

Tabla 2. Ejemplo de matriz de calificaciones (en una escala de 1-5).

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Usuario A	4	?	3	5
Usuario B	?	5	4	?
Usuario C	5	4	2	?
Usuario D	2	4	?	1

se aplicó el Coeficiente de Pearson, debido a que se utiliza el enfoque basado en el ítem del Filtro Colaborativo. Después se eligen los vecinos más cercanos para el ítem objetivo, con todos los vecinos (k) se predice la calificación del ítem y por último se genera una lista previa de las recomendaciones con el algoritmo Top-N [6].

En caso de que el usuario no tenga aún restaurantes calificados entonces se emplea el Algoritmo del Filtro Basado en Contenido. En la Fig. 3b se presenta el proceso del algoritmo que genera una lista previa de recomendaciones.

Como objetivo, el algoritmo Basado en Contenido debe seleccionar y determinar los ítems basados en la correlación y relación entre el contenido de cada ítem (en este trabajo son los restaurantes) y los gustos de los usuarios. Este algoritmo en particular utiliza técnicas de Minería de Texto: Recuperación de Información, Extracción de Información, Categorización y PLN (Procesamiento de Lenguaje Natural) [28].

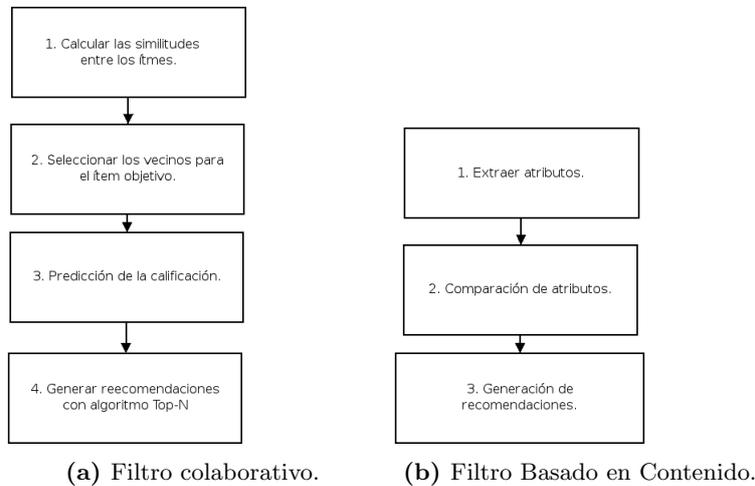


Fig. 3. Procesos de cada algoritmo del sistema de recomendación.

Para iniciar el proceso del Filtro Basado en Contenido, se obtienen los comentarios positivos que los usuarios han generado para cada restaurante y los gustos del usuario que desea obtener recomendaciones. La polaridad de los comentarios se mide con un Clasificador Bayesiano [25]. Después mediante técnicas de minería de texto, se extraen los atributos de los ítems. Posteriormente se comparan los atributos con los gustos del usuario y finalmente se generan las recomendaciones.

Este Sistema de Recomendación aplica un filtro más a las recomendaciones generadas. Con la ubicación del usuario se eligen los restaurantes mas cercanos al usuario. Con esto termina el flujo del sistema, generando una lista de restaurantes recomendados para el usuario.

4. Evaluación experimental

La evaluación de los algoritmos integrados en el Sistema de Recomendación Híbrido que genera recomendaciones de restaurantes para los usuarios, hace uso de una base de datos que fue poblada con datos generados en TripAdvisor y Foursquare. En la Tabla 3 se describen los datos usados en las pruebas.

4.1. Precisión y recall

Para el análisis de la información resultante del SR Híbrido, se utilizaron dos métricas de evaluación sobre las recomendaciones obtenidas como lo son: Precisión y Recall que se enlistan a continuación en las ecuaciones 1 y 2.

La Precisión es la fracción de elementos recomendados que son realmente relevantes para el usuario [15]:

$$Precisión = \frac{Items\ correctamente\ recomendados}{Total\ de\ items\ recomendados}. \quad (1)$$

Tabla 3. Datos que se utilizan en las pruebas.

Dato
685 lugares como restaurantes, bares, cafeterías, entre otros.
4764 comentarios relacionados a los lugares.
65274 calificaciones otorgadas a los lugares.
53051 usuarios que interactúan con los lugares.
14876 comentarios usados en el clasificador Bayesiano.

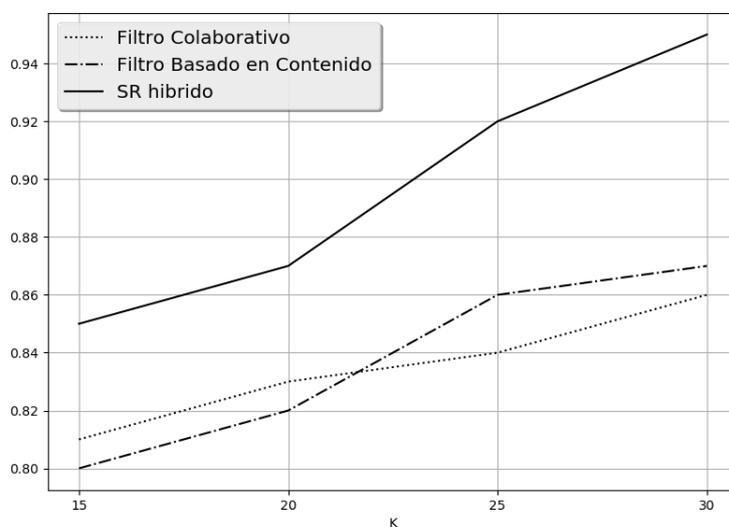


Fig. 4. Comparación de la Precisión entre Filtro Colaborativo, Filtro Basado en Contenido y SR Híbrido.

Recall se puede definir como la fracción de elementos relevantes que también forman parte del conjunto de elementos recomendados [15]:

$$Recall = \frac{Items\ correctamente\ recomendados}{Total\ de\ items\ recomendados\ útiles} \quad (2)$$

4.2. Resultados de la evaluación

Después calcular los valores de Precisión y Recall para cada usuario, se obtuvo la media de estas métricas para el sistema empleando solo el Filtro Colaborativo, posteriormente se aplicó lo mismo al Filtro Basado en Contenido y al final el cálculo se realizó para el Sistema de Recomendación Híbrido. Las Fig. 4 y 5 muestran una comparativa de los resultados para Precisión y Recall de las diferentes pruebas que se realizaron con el Sistema de Recomendación.

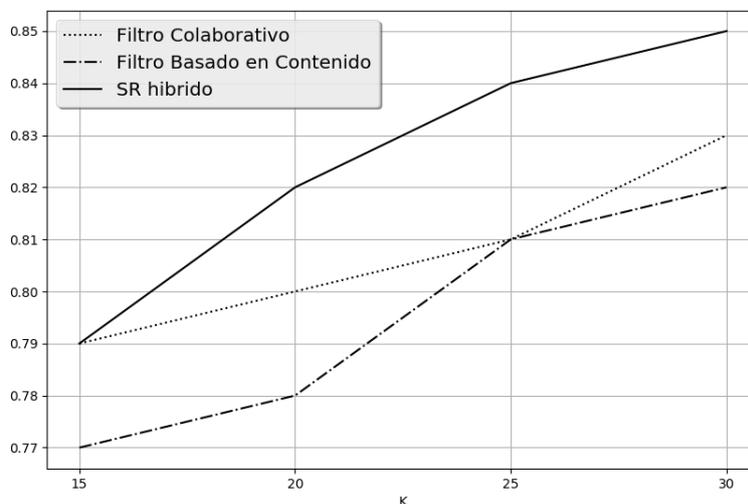


Fig. 5. Comparación de Recall entre Filtro Colaborativo, Filtro Basado en Contenido y SR Híbrido.

Como se puede apreciar, los resultados presentados en la gráfica de la Precisión aumentan según el incremento del valor de los vecinos (k) para todos los casos. La probabilidad de que un ítem recomendado por el sistema sea relevante para el usuario, supera el 80% para las 4 variantes probadas en el caso del Filtro Colaborativo y Basado en Contenido y se mantiene sobre el 85% cuando se utiliza el SR Híbrido. En el caso de Recall, la probabilidad de que un ítem relevante sea recomendado, también crece en relación directa con el valor de k y esta probabilidad se mantiene por encima del 77% cuando se evalúa el Filtro Colaborativo y Basado en Contenido, y sobre el 79% para la evaluación del SR Híbrido.

Con los resultados presentados se observa que el Sistema de Recomendación Híbrido mejora cuando exista la combinación de algoritmos, ya que encontramos algunas deficiencias en el uso particular de los algoritmos.

5. Conclusiones y trabajos a futuro

En este trabajo se ha presentado un Sistema de Recomendación que genera una lista de restaurantes recomendados para los consumidores basándose en las preferencias y localización del consumidor. La problemática se resolvió aplicando técnicas de Minería de Texto y Minería de Datos, usando los algoritmos de Filtro Colaborativo y Basado en Contenido.

Entre los objetivos que se plantearon al inicio del trabajo fue por un lado, obtener y construir un conjunto de datos con la temática de restaurantes. Y por otro lado, evaluar la efectividad del Sistema de Recomendación Híbrido con

los diferentes casos que se puedan presentar al atender las necesidades de los consumidores.

Los resultados obtenidos mostraron que la técnica híbrida brinda mejores resultados en comparación si los algoritmos de Filtro Colaborativo y Basado en Contenido trabajaran individualmente. Las desventajas que presentan estos algoritmos se mitigan al fusionarlos y optar por la técnica híbrida. Con esto se comprueba que la combinación de algoritmos proporciona una lista de recomendaciones acorde al usuario, donde es importante enfatizar que la Minería de Texto y la Minería de Datos son áreas de investigación que tienen un amplio sector de aplicación.

Como trabajo futuro se propone aplicar una técnica de desarrollo de software para implementar el Sistema de Recomendación Híbrido. Para tal situación es conveniente desarrollar un servicio de web para conectar el sistema con los dispositivos móviles, y así tener una ventaja más en el sistema, ya que con esto se tiene como ventaja que el consumidor tenga las recomendaciones de los restaurantes en su dispositivo móvil.

Referencias

1. Amatriain, X.: Recommender systems (machine learning summer school 2014 @ cmu) (Febrero 2018), <https://www.slideshare.net/xamat/recommender-systems-machine-learning-summer-school-2014-cmu/>
2. Ashraf, A., Rabbani, I.M., Martínez-Enriquez, A.M., Muhammad, A.: A user-centered approach for cloud service selection and recommendation. *Research in Computing Science* 130, 99–110 (Diciembre 2016)
3. Bogers, T., Van den Bosch, A.: Collaborative and content-based filtering for item recommendation on social bookmarking websites. Submitted to CIKM 9 (2009)
4. Burke, R.: The adaptive web. chap. Hybrid Web Recommender Systems, pp. 377–408. Springer-Verlag, Berlin, Heidelberg (2007), <http://dl.acm.org/citation.cfm?id=1768197.1768211>
5. d. C. L. Soares, J., Suyoto, Santoso, A.J.: M-guide: Hybrid recommender system tourism in east-timor. In: 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT). pp. 303–309 (Sept 2017)
6. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)* 22(1), 143–177 (2004)
7. Foursquare, Inc.: Acerca de nosotros (Febrero 2018), <https://es.foursquare.com/about/>
8. Foursquare, Inc.: Venue categories (Febrero 2018), <https://developer.foursquare.com/docs/resources/categories>
9. Fundación Wikimedia, Inc.: Sistema de posicionamiento global (Febrero 2018), https://es.wikipedia.org/wiki/Sistema_de_posicionamiento_global
10. García, J.R.A., García, J.V.H., Tello, E.V., de la Rosa, G.R., Sánchez, C.S.: Sistema de recomendación de música basado en aprendizaje semi-supervisado. *Research in Computing Science* 94, 97–109 (Mayo 2015)
11. González, O.E., Jacques, S.M.: Estado del arte en los sistemas de recomendación. *Research in Computing Science* 135, 25–40 (Noviembre 2017)
12. Guzmán-Luna, J., Torres, I.D., Vallejo, S.: Un sistema recomendador móvil de rutinas de ejercicio basado en el perfil del usuario. *Research in Computing Science* 94, 137–149 (Mayo 2015)

13. Habib, M.A., Rakib, M.A., Hasan, M.A.: Location, time, and preference aware restaurant recommendation method. In: 2016 19th International Conference on Computer and Information Technology (ICCIT). pp. 315–320 (Dec 2016)
14. Instituto Nacional de Estadística y Geografía: Censos económicos 2014. resultados definitivos (Febrero 2018), <http://www.inegi.org.mx/est/contenidos/proyectos/ce/ce2014/default.aspx>
15. Isinkaye, F., Folaajimi, Y., Ojokoh, B.: Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal* 16(3), 261–273 (2015), <http://www.sciencedirect.com/science/article/pii/S1110866515000341>
16. Jain, S., Grover, A., Thakur, P.S., Choudhary, S.K.: Trends, problems and solutions of recommender system. In: International Conference on Computing, Communication Automation. pp. 955–958 (Mayo 2015)
17. Ketmaneechairat, H., Kongketwanich, C., Naijit, T.: Recommender system for thai food cooking on smartphone. In: 2017 Twelfth International Conference on Digital Information Management (ICDIM). pp. 169–178 (Sept 2017)
18. Mansur, F., Patel, V., Patel, M.: A review on recommender systems. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). pp. 1–6 (Marzo 2017)
19. Martí, M.: Qué es el web scraping? introducción y herramientas (Febrero 2018), <https://sitelabs.es/web-scraping-introduccion-y-herramientas>
20. Martínez, L., Rodríguez, R.M., Espinilla, M.: Reja: A georeferenced hybrid recommender system for restaurants. In: 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. vol. 3, pp. 187–190 (Sept 2009)
21. Otero, J.D.C., Gonzalez1, M.J.R., Franco, M.T.M.: Esbozo de una técnica para la recomendación de productos turísticos en cuba usando gis. *Research in Computing Science* 79, 21–36 (Octubre 2014)
22. Pal, A., Parhi, P., Aggarwal, M.: An improved content based collaborative filtering algorithm for movie recommendations. In: 2017 Tenth International Conference on Contemporary Computing (IC3). pp. 1–3 (Agosto 2017)
23. Patel, B., Desai, P., Panchal, U.: Methods of recommender system: A review. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). pp. 1–4 (Marzo 2017)
24. Renjith, S., Anjali, C.: A personalized mobile travel recommender system using hybrid algorithm. In: 2014 First International Conference on Computational Systems and Communications (ICCSC). pp. 12–17 (Diciembre 2014)
25. Sanzón, Y.M., Vilariño, D., Somodevilla, M.J., Zepeda, C., Tovar, M.: Modelos para detectar la polaridad de los mensajes en redes sociales. In: *Research in Computing Science*
26. Stabb, S., Werther, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D.R., Paris, C., Knoblock, C.: Intelligent systems for tourism. *IEEE Intelligent Systems* 17(6), 53–66 (Noviembre 2002)
27. TripAdvisor, Inc.: Acerca de tripadvisor (Febrero 2018), <https://tripadvisor.mediaroom.com/mx-about-us>
28. Vijayarani, S., Ilamathi, M.J., Nithya, M.: Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks* 5(1), 7–16 (2015)
29. Wijaya, K.A., Suyoto, Santoso, A.J.: M-guide: Recommending systems of food centre in buleleng regency. In: 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIT). pp. 310–314 (Sept 2017)

Saúl Pérez, Mary Carmen Cuecuecha, José Federico Ramírez, José Crispín Hernández

30. Yang, Z., Wu, B., Zheng, K., Wang, X., Lei, L.: A survey of collaborative filtering-based recommender systems for mobile internet applications. *IEEE Access* 4, 3273–3287 (Mayo 2016)