B U A P

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Un enfoque híbrido para sistemas de recomendación aplicado a un dominio de e-commerce

Tesis para obtener el título de:

Doctor en Ingeniería del Lenguaje y del Conocimiento

Presenta: Victor Giovanni Morales Murillo

Asesores: Dr. David Eduardo Pinto Avendaño

Dr. Franco Rojas López

Puebla, Puebla, México

Agosto 2024

Abstract

In the digital era, recommendation systems are one of the main applications of data science and artificial intelligence for humans because they have become integrated into our daily lives and are powerful tools for making informed, effective, and efficient decisions. This technology addresses the information overload, a growing problem in a globalized world where more than 60% of the population are internet users who navigate an average of 6 hours daily in cyberspace. E-commerce is an activity strongly driven by the nature of the digital era and recently by the pandemic caused by the SARS-CoV-2 virus. The COVID-19 pandemic caused multiple countries to accelerate their digitization processes and shift their sales channels to e-commerce, as reported by the Mexican Online Sales Association in the case of Mexico. This doctoral thesis conducted applied and experimental research with a quantitative approach on hybrid recommendation systems for e-commerce. This research comprises four relevant publications in the area of recommendation systems. The first publication is about a systematic literature review on hybrid recommendation approaches to identify their main trends in challenges, methodologies, application domains, datasets, evaluation methods, and areas of opportunity. The second publication is about a hybrid recommendation model based on information retrieval for Mexican tourist texts from Rest-Mex 2022, in which a hybridization technique by selection was applied on content-based filtering and collaborative filtering through a vector space model developed with a TF-IDF weighting scheme and applying cosine similarity. This model achieved the second and third-best results in the ranking of the recommendation system task of Rest-Mex 2022 and presented a minimal difference with the first place. In the third publication, the sentiment analysis task of Rest-Mex 2023 was addressed because this type of technology can provide feedback to a recommendation system to improve the quality of recommendations. In this research, the polarity, type, and country of the tourist reviews were classified using a domain adaptation based on the RoBERTa transformer, and different training strategies were generated, achieving first place in the ranking of Rest-Mex 2023. Finally, the fourth publication analyzed session-based recommendation systems for e-commerce using the latest advances in NLP, such as transformer architectures, to account for users' dynamic and short-term preferences. 102 million Amazon reviews were pre-processed to generate two datasets (one domain-specific and the other multidomain), and it was verified that the XLNet transformer with different training strategies based on language modeling outperformed a GRU RNN.

Resumen

En la era digital los sistemas de recomendación son una de las principales aplicaciones de ciencia de datos e inteligencia artificial para el ser humano debido a que se han integrado en nuestra vida diaria y son poderosas herramientas para la toma de decisiones informadas, efectivas y eficientes. Esta tecnología aborda la sobrecarga de información que es un problema que crece cada día en un mundo globalizado en donde más del 60% de la población es un usuario de internet que navega en promedio 6 horas diarias en el ciberespacio. El e-commerce es una actividad que es fuertemente impulsada por la naturaleza de la era digital y recientemente por la pandemia causada por el virus SARS-CoV-2. La pandemia de COVID-19 provoco que múltiples países aceleraran sus procesos de digitalización y cambiaran sus canales de venta al comercio electrónico como fue el caso en México reportado por la Asociación Mexicana de Venta Online. En esta tesis doctoral se realizó una investigación aplicada y experimental con un enfoque cuantitativo con el objetivo de generar y evaluar un sistema de recomendación híbrido para el e-commerce. Esta investigación está compuesta por cuatro publicaciones relevantes sobre el área de sistemas de recomendación. La primera es sobre una revisión sistemática de la literatura sobre los enfoques híbridos de recomendación para identificar sus principales tendencias en desafíos, metodologías, dominios de aplicación, conjuntos de datos, formas de evaluación y áreas de oportunidad. La segunda publicación es sobre un modelo híbrido de recomendación basado en recuperación de información para textos turísticos mexicanos del Rest-Mex 2022, en el cual se aplicó una técnica de hibridación por selección sobre el filtrado basado en contenido y el filtrado colaborativo a través de un modelo espacio vectorial desarrollado con un esquema de ponderación TF-IDF y aplicando la similitud del coseno. Este modelo logró el segundo y tercer mejor resultado en el ranking de la tarea de sistema de recomendación del Rest-Mex 2022 y presentó una mínima diferencia con el primer lugar. En la tercera publicación se abordó la tarea de análisis de sentimiento del Rest-Mex 2023 debido a que este tipo de tecnología puede retroalimentar a un sistema de recomendación para mejorar la calidad de las recomendaciones, en esta investigación se clasificó la polaridad, el tipo y el país de las reseñas turísticas utilizando una adaptación de dominio basada en el transformer RoBERTa y se generaron diferentes estrategias de entrenamiento, con las cuales, se logró el primer lugar en el ranking del Rest-Mex 2023. Finalmente, en la cuarta publicación se analizaron los sistemas de recomendación basados en sesiones para el e-commerce utilizando los últimos avances de NLP como son las arquitecturas transformer para tomar en cuenta las preferencias dinámicas y a corto plazo de los usuarios, 102 millones de reseñas de Amazon fueron pre-procesadas para generar dos conjuntos datos (de un dominio y otro multidominio) y se comprobó que el transformer XLNet con diferentes estrategias de entrenamiento basadas en el modelado del lenguaje superó a una RNN GRU.

Agradecimientos

A Dios y mi familia por su apoyo constante e incondicional durante toda mi vida.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por el apoyó otorgado a través de la beca nacional al CVU 903146.

Así como a la Benemérita Universidad Autónoma de Puebla (BUAP), la Vicerrectoría de investigación y Estudios de Posgrado (VIEP), la Dirección de Innovación y Transferencia de Conocimiento (DITCO), la Facultad de Ciencias de la Computación (FCC) y el laboratorio de Ingeniería de Lenguaje y del Conocimiento (LKE), al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM) y a la School of Enterprise Computing and Digital Transformation de la Technological University Dublin (TU Dublin), Tallaght Campus por todas las facilidades y apoyos otorgados durante mis estudios de doctorado.

Al vicerrector de la BUAP el Dr. Ygnacio Martínez Laguna, al director de la DITCO el Dr. David Pinto, a la directora de la FCC la M.I. María del Consuelo Molina García y a la secretaria de posgrado la Dra. María Josefa Somodevilla García por su apoyo para gestionar el financiamiento para este proyecto de tesis doctoral.

A mis asesores el Dr. David Pinto y el Dr. Franco Rojas quienes con su conocimiento, experiencia y buen carácter me acompañaron a lo largo de mis estudios y me prepararon para alcanzar esta meta.

A mis sinodales, la Dra. María Josefa Somodevilla García, el Dr. Fernando Pérez Téllez y el Dr. Rafael Guzmán Cabrera por sus observaciones y comentarios.

A la Dra. Helenas Gómez Adorno del IIMAS de la UNAM y al Dr. Fernando Pérez Téllez de la TU Dublin por supervisarme durante mis estancias de investigación y enriquecer mi tesis doctoral.

Finalmente, pero no menos importante, a mis profesores, amigos, compañeros y todos aquellos que creyeron en mí apoyándome y animándome.

Dedicatoria

Para mi hijo hermoso Victor Caleb y mi hermosa esposa Damaris quienes siempre me han amado, apoyado y motivado incondicionalmente para alcanzar este sueño.

A mi madre Adoralida y mi padre Victor Baruch por darme la oportunidad de disfrutar cada momento de mi vida como el de convertirme en doctor.

A mi hermano Rodolfo Baruch y su familia por su apoyo incondicional.

A mis abuelos y en especial a mi abuelo el ex. Secretario General de los Ingenios Azucareros y ex. Dip. Manuel Antonio Morales Álvarez por ser una de mis principales fuentes de motivación, admiración y respeto.

¡A toda mi familia que siempre creyó en mí!

A mi asesor el Dr. David Pinto que más que un extraordinario asesor, lo considero un gran amigo que me ha motivado y guiado desde la licenciatura hasta el doctorado.

A todos ellos dedico esta tesis, pues son el ingrediente principal de esta investigación.

Contenido

Lista de figuras

Lista de tablas

Capítulo1: Introducción

Los sistemas de recomendación surgen por el exceso de información que existe en el Internet, el cual limita la búsqueda de información relevante para el usuario. Estos sistemas son herramientas que mejoran la toma de decisiones y la experiencia de navegación de sus usuarios lo que representa un gran beneficio para las compañías que los utilizan porque aumentan sus ventas y reducen la tasa de abandono de sus usuarios llamado en inglés churn rate [1]. Por tal motivo, en la actualidad las principales compañías de tecnología y del e-commerce utilizan los sistemas de recomendación dentro de sus principales servicios, por ejemplo, Netflix reportó que al menos el 75% de sus rentas y descargas provenían de su sistema de recomendación por lo que realizó el famoso premio de Netflix para optimizar sus algoritmos. Otro ejemplo es el de Amazon que desarrollo su propia técnica de recomendación con base en el filtrado colaborativo pero basada en la similitud de los productos o ítems en lugar de usar la similitud entre usuarios, el filtrado colaborativo basado en el ítem le permitió a Amazon dirigir marketing automatizado y personalizado para sus usuarios que incrementó sus ventas y su tasa de clics [2]. También, Spotify es una de las principales compañías de transmisión de música que compite con grandes corporaciones como Apple, Google y Amazon gracias a su sistema de recomendación híbrido que combina el procesamiento del lenguaje natural, el filtrado colaborativo y el filtrado basado en contenido [3]. Finalmente, en las redes sociales como Facebook y Twitter, los sistemas de recomendación son fundamentales para abordar el problema de sobrecarga de información para sus usuarios y lanzar campañas de marketing digital gracias a la rica fuente de información que proporcionan sus usuarios a través de estas redes sociales.

El comercio electrónico o e-commerce es uno de los principales canales de venta para las empresas porque les facilita a sus usuarios realizar compras desde la comodidad de sus hogares sin la necesidad de ir a una tienda física y gastar parte de su tiempo. En México, el Instituto Nacional de Estadística y Geografía (INEGI) reporto en un censo del año 2019 [4] que existen 4.9 millones de establecimientos privados y paraestatales, de los cuales 4.89 millones representan a las micro, pequeñas y medianas empresas (MiPyMEs) y la Asociación Mexicana de Venta Online (AMVO) reporto que 6 de cada

10 pequeñas y medianas empresas (PyMEs) venden en línea [5]. A pesar de que el e-commerce en México se encuentra en una etapa de apertura, muchas empresas tuvieron que migrar al comercio electrónico a causa de la pandemia ocasionada por el virus SARS-CoV-2 y se espera para finales del año 2021 que el 35% de las ventas totales de las PyMEs sea por el canal del e-commerce. Por tal motivo, en esta tesis se propone generar un enfoque híbrido para sistemas de recomendación en un dominio de e-commerce para beneficiar e impulsar al comercio electrónico.

1.1 Antecedentes

El antecedente de este trabajo de investigación se desarrolló en la tesis de maestría titulada: "*Un nuevo paradigma basado en recuperación de información para sistemas de recomendación*" [6], en la cual se propusieron dos paradigmas basados en recuperación de información para sistemas de recomendación, en los cuales se experimentó en el cálculo de la similitud entre usuarios e ítems; se implementó el modelo de espacio vectorial con el algoritmo Cosine Score y se utilizaron de forma individual las técnicas de recomendación de filtrado colaborativo basado en el usuario y filtrado colaborativo basado en el ítem. Además, se evaluaron ambas técnicas de recomendación con un conjunto de datos estándar del dominio de películas llamado MovieLens. Los resultados obtenidos en la anteriormente mencionada tesis de maestría alientan el desarrollo de métodos avanzados para mejorar el rendimiento de sistemas de recomendación. Por esta razón, en esta propuesta de proyecto doctoral se propone un enfoque híbrido, en el cual se pretende considerar los elementos de mayor valor obtenidos en el trabajo previo.

1.2 Justificación

Los sistemas de recomendación representan un alto impacto social, económico y tecnológico a nivel internacional porque son herramientas que mejoran la toma de decisiones para sus usuarios. Estos sistemas son fundamentales para grandes compañías que invierten grandes cantidades de dinero buscando optimizar sus algoritmos de recomendación [2] y son una de las principales áreas de investigación [7]. Esta investigación propone generar un enfoque híbrido mediante la combinación de dos técnicas de recomendación con la finalidad de mejorar la eficiencia global en comparación de aquellas que utilizan las técnicas de forma individual. Además, este enfoque híbrido servirá como una alternativa para minimizar el problema de escalabilidad. Finalmente, este proyecto beneficiará las investigaciones futuras relacionadas con sistemas de recomendación con enfoques híbridos, así como también

aquellas aplicaciones que usen este tipo de técnicas, por ejemplo, proyectos que impacten en la comunidad universitaria de la BUAP y en el e-commerce de México.

1.3 Planteamiento del problema

En el estudio titulado *Digital 2020: Global digital overview* [8] se presenta que aproximadamente 4.5 mil millones de personas, que representan el 60% de la población, usan el Internet. El usuario promedio navega en Internet 6 horas y 43 minutos, lo que representa más del 40% del tiempo en el que permanece despierta una persona; y existen más de 3.8 mil millones de usuarios en redes sociales. Estas estadísticas son un gran impulso para el comercio electrónico. Pero, además muy importante para este trabajo de investigación es que las principales compañías de tecnología como Google, Netflix, Facebook, Amazon y Spotify utilizan los sistemas de recomendación dentro de sus principales servicios e invierten grandes cantidades de dinero para optimizar sus algoritmos de recomendación [2]. El estudio de los sistemas de recomendación es una de las principales áreas de investigación, la cual es relativamente reciente en combinación con el uso de métodos de aprendizaje automático [7].

Los sistemas de recomendación surgen por el exceso de información en el Internet, que limita la búsqueda de información importante para el usuario [9]. Los sistemas de recomendación son herramientas que sugieren una amplia gama de ítems relevantes para los usuarios, como pueden ser productos, películas, canciones, amigos, alimentos, ropa, restaurantes, páginas web, noticias y otros ítems. Las recomendaciones se pueden realizar con base en las características del usuario o del ítem; estas características dependen de las principales técnicas de recomendación como son el filtrado basado en contenido, el filtrado colaborativo y el filtrado híbrido.

Los principales problemas para los sistemas de recomendación son: (1) la precisión: que se refiere al grado de eficiencia de las recomendaciones (2) la escalabilidad: que requiere procesar altos volúmenes de información con mínimos tiempos de respuesta, (3) el arranque en frio: que requiere una cantidad suficiente de usuarios e ítems para buscar usuarios o ítems similares, (4) la escasez de datos: requiere que la mayoría de los usuarios califiquen la mayoría de los ítems, sin embargo, el alto número de usuarios e ítems dificulta más el problema, (5) la sobre-especialización: requiere que las recomendaciones no se basen únicamente en el perfil del usuario para que se puedan sugerir nuevos ítems u otras opciones, (6) la diversidad: requiere que los ítems recomendados sean de diferentes grupos o clases, (7) la novedad: requiere que las recomendaciones contengan otros ítems nuevos, (8) la serendipia: requiere no solamente novedad, sino también que las recomendaciones sorprendan al usuario, (9) la privacidad: es un desafío importante porque los sistemas de recomendación deben conocer información privada de los usuarios para mejorar sus recomendaciones, sin

embargo, es complejo obtener la información privada, (10) la oveja gris: se presenta cuando un usuario no tiene usuarios similares, es decir, el usuario no pertenece a un grupo [7].

En el laboratorio de Ingeniería del Lenguaje y del Conocimiento dirigido por el Dr. David Eduardo Pinto Avendaño de la Benemérita Universidad Autónoma de Puebla (BUAP) se tiene el interés por generar un enfoque híbrido para sistemas de recomendación aplicado a un dominio de e-commerce para abordar los problemas de precisión y escalabilidad en las recomendaciones. Además, este proyecto surge por el interés de la Benemérita Universidad Autónoma de Puebla para recomendar sus productos y servicios, sin embargo, esta investigación también beneficiará a otros proyectos que impulsen al e-commerce.

1.4 Objetivos

A continuación, se presenta el objetivo general, los objetivos específicos y las preguntas de investigación.

1.4.1 Objetivo general

Generar un enfoque híbrido para sistemas de recomendación aplicado a un dominio de e-commerce.

1.4.2 Objetivos específicos

- Proponer dos paradigmas basados en técnicas individuales de recomendación para evaluarse e hibridarse.

- Implementar un modelo de recomendación híbrido con los paradigmas propuestos para impactar en la precisión y escalabilidad de sus recomendaciones en comparación de utilizar los paradigmas propuestos de forma individual.

- Evaluar con un conjunto de datos estándar los paradigmas propuestos de forma individual y de forma conjunta con el modelo de recomendación híbrido para comparar los resultados.

1.4.3 Preguntas de investigación

- ¿Cómo mejorar la precisión de las recomendaciones con un modelo de recomendación híbrido en comparación de utilizar una sola técnica de recomendación?

- ¿Cómo minimizar el problema de escalabilidad con un modelo de recomendación híbrido en comparación de utilizar una sola técnica de recomendación?

1.5 Hipótesis

Las hipótesis de investigación son causales multivariadas [10] porque se presenta como variable independiente la implementación de la hibridación de dos técnicas de recomendación y como variables dependientes la precisión de las recomendaciones y la minimización del problema de escalabilidad. Las hipótesis de investigación son las siguientes.

H_1: *La implementación de la hibridación de dos técnicas de recomendación mejorará la precisión de las recomendaciones, en comparación con las técnicas que implementan de forma individual las técnicas de recomendación.*

H_2: *La implementación de la hibridación de dos técnicas de recomendación minimizará el problema de escalabilidad, en comparación con las técnicas que implementan de forma individual las técnicas de recomendación.*

1.6 Alcances y limitaciones

Se realizará una investigación aplicada y experimental con un enfoque cuantitativo ya que se efectuarán diferentes tipos de pruebas experimentales que serán evaluadas cuantitativamente usando fórmulas clásicas de medición de rendimiento, sobre un conjunto de datos con la finalidad de determinar la calidad de las propuestas hechas en un enfoque híbrido para un sistema de recomendación. Además, la investigación tiene diferentes alcances metodológicos que permiten explorar las técnicas hibridas de los sistemas de recomendación, describir los sistemas de recomendación, correlacionar las diferentes variables que pueden manipularse en los sistemas de recomendación y explicar la causa-efecto sobre el rendimiento del sistema de recomendación. Las referencias básicas sobre metodología de la investigación que se utilizan son la de [10] y la de [11], las cuales se utilizaron en las materias de proyectos de investigación I y II.

1.7 Organización de la tesis

La organización de la tesis es la siguiente. En el *capítulo 1*, se presenta la introducción que contiene los antecedentes, la justificación, el problema de investigación, los objetivos, la hipótesis y el alcance. En el *capítulo 2*, se presenta el marco teórico que contiene las técnicas de recuperación de información, los enfoques de recomendación y las métricas de evaluación. En el *capítulo 3*, se presenta la revisión sistemática del estado del arte sobre enfoques híbridos de recomendación. En el *capítulo 4*, se presenta un modelo recomendación híbrido basado en RI para textos turísticos mexicanos. En el *capítulo 5*, se presenta un sistema de recomendación multidominio basado en Transformer para el e-commerce. Finalmente, en el último capitulo se presentan las conclusiones y trabajos futuros.

Capítulo 2: Marco teórico

El marco teórico se genera con una primera revisión de la literatura con la metodología de la heurística y la hermenéutica [12], en la heurística se realiza la búsqueda, clasificación y compilación de las principales fuentes de información sobre sistemas de recomendación, en la cual se identificaron 22 publicaciones relevantes y en el método de interpretación de textos (hermenéutica) se analizan 19 publicaciones de las 22 publicaciones previamente identificadas, de las cuales 12 investigaciones son descriptivas y 8 investigaciones son experimentales, estas publicaciones se muestran en la Tabla 1. El marco teórico se construye utilizando el método por índices [10] con base en la previa revisión de la literatura sobre sistemas de recomendación con enfoques híbridos, este método consiste en definir un índice general que es continuamente refinado hasta ser sumamente especifico y después se coloca la información en el lugar correspondiente dentro del esquema.

Tabla 1. Publicaciones analizadas para generar el marco teórico.

#	Cita	Tema	Tipo de investigación
1	[13]	Evaluación del filtrado colaborativo de sistemas de recomendación.	Descriptiva.
2	[14]	Estudio sobre sistemas de recomendación para nuevos datos.	Descriptiva y experimental.
3	[15]	Sistema de recomendación mediante recuperación de información.	Experimental.
4	[16]	Sistema de recomendación híbrido: usando perfiles, palabras clave y calificaciones.	Experimental.
5	[17]	Sistemas de recomendación: Principios, métodos y evaluación.	Descriptiva.
6	[2]	Sistemas de recomendación más que una competencia de matrices.	Descriptiva.

7	[9]	Sistema ecléctico de filtrado de información basado en inteligencia computacional para recomendación de atractivos turísticos del caribe colombiano.	Experimental y aplicada.
8	[18]	Sistemas de recomendación basados en inteligencia artificial.	Descriptiva.
9	[19]	Estado del arte en los sistemas de recomendación.	Descriptiva.
10	[3]	Filtrado colaborativo de Spotify.	Descriptiva.
11	[7]	Estudio sobre desafíos y soluciones para sistemas de recomendación.	Descriptiva.
12	[20]	Sistemas de Recomendación: un enfoque a las técnicas de filtrado.	Descriptiva.
13	[21]	Una revisión de tendencias y técnicas en sistemas de recomendación.	Descriptiva.
14	[22]	Sistema de recomendación híbrido basado en ubicación móvil de e-tourism con evaluación de contexto.	Experimental.
15	[23]	Un estudio sobre el sistema de recomendación de comercio electrónico basado en Big Data.	Experimental.
16	[24]	Un nuevo sistema de recomendación híbrido mejorado para mitigar los problemas de arranque en frío.	Experimental.
17	[25]	Un estudio comparativo sobre enfoques de recomendación.	Descriptiva y comparativa.
18	[26]	Un algoritmo de recomendación sobre currículum vitae basado en K-means++ y Part-of-speech tf-idf.	Experimental.
19	[27]	Una revisión de sistemas de recomendación: Limitaciones, estudios y desafíos.	Descriptiva.

2.1 Sistemas de recomendación

2.1.1 Definición

A continuación, se presentan diferentes definiciones sobre los sistemas de recomendación para la conceptualización del término. Estas definiciones son similares y sus diferencias se deben a los diversos enfoques de recomendación. En el trabajo de [20] los sistemas de recomendación se encargan de sugerir ítems relacionados a las preferencias de sus usuarios entre una gran cantidad de opciones disponibles; además, estos sistemas están compuestos por datos de contexto (información inicial del sistema antes de la recomendación), datos de entrada (información que el usuario comunica con el sistema antes de la recomendación) y algoritmos de recomendación (algoritmos que combinan los datos de contexto con los datos de entrada para generar recomendaciones). En el trabajo de [25] los sistemas de recomendación están destinados a ayudar a los usuarios para descubrir ítems relevantes entre varios ítems debido a los altos volúmenes de información en el Internet. En el trabajo de [27] se define como una herramienta de información que ayuda a los usuarios a buscar ítems preferidos entre un gran conjunto de ítems y el principal objetivo es predecir la calificación que un usuario le asignará a un ítem.

El sistema de recomendación usa una función de utilidad que estima el interés que tiene un usuario por sus ítems recomendables. Su definición formal es la siguiente: Sea U un conjunto de todos los usuarios e I el conjunto de todos los posibles ítems recomendables. Definimos $\underline{r}$ como una función que estima la utilidad de un ítem i para el usuario u, es decir:

$$\underline{r}: U \times I \rightarrow R, \qquad \textit{donde } R \textit{ es un conjunto ordenado.}$$

Generalmente se crea una lista ordenada de ítems basándose en el interés estimado y se seleccionan los n mejores ítems [19].

2.1.2 Historia

Los sistemas de recomendación se originan por el exceso de información expuesta en diferentes ámbitos, que limita a los usuarios a determinar los ítems que son relevantes para ellos [20]. Los campos que han contribuido en esta área de investigación son los sistemas de información, la recuperación de información, el aprendizaje automático, la interacción humano-computadora, el marketing y la física.

En la década de 1960, surge el enfoque basado en contenido con la idea de usar una computadora para filtrar información de acuerdo con las preferencias del usuario. Los primeros sistemas empleaban palabras claves explícitas proporcionadas por los usuarios para clasificar y filtrar documentos, como la ponderación de coincidencias de palabras claves. Después se aplicaron técnicas más elaboradas como los vectores de términos ponderados como tf-idf o métodos de análisis de documentos como la indexación semántica latente (LSI). En la década de 1980, surge el enfoque colaborativo con el problema de los correos electrónicos basura expuesto por el entonces presidente de ACM, el sistema de filtrado de correos llamado Tapestry de la compañía de Xerox PARC introdujo la idea que algunos lectores podían calificar mensajes y otros lectores podían tener acceso a esos mensajes, sin embargo, fue hasta la década de 1990 cuando el sistema GroupLens introduce la idea del sistema Better Bit Bureau, el cual hacía predicciones automatizadas sobre qué ítems le gustarían a los usuarios basándose en un esquema de vecinos cercanos, iniciando así las aportaciones empíricas al enfoque colaborativo. A inicios de la década del 2000, surgen varias aplicaciones de sistemas de recomendación para el comercio electrónico y en el año 2003 Amazon desarrolla un enfoque colaborativo ítem-ítem que representa un alto impacto para su marketing, ya que este enfoque abordaba el problema de la escalabilidad y las necesidades de realizar recomendaciones en tiempo real. A mediados de la década del 2000, se formula el problema de los sistemas de recomendación como un problema de completar una matriz de calificaciones; en esa matriz de calificaciones, las filas representan a los usuarios, las columnas representan a los ítems y las celdas contienen las calificaciones de los usuarios hacia los ítems preferidos en una escala de 1 a 5 estrellas o en función del comportamiento del usuario. La solución a este problema consiste en predecir las celdas no calificadas por el usuario, por tal motivo, se introdujeron técnicas de aprendizaje automático como son de agrupamiento, clasificación, regresión o descomposición de valores. En el año 2009, Netflix otorgó un millón de dólares como premio al ganador del concurso Netflix Price; este concurso consistía en proponer un algoritmo de recomendación que superara en un 10% de RMSE el rendimiento de los algoritmos de Netflix. En este concurso, las técnicas de aprendizaje automático y factorización de matrices demostraron ser exitosos, sin embargo, los sistemas de recomendación están lejos de ser un problema resuelto porque la predicción de calificaciones puede ser insuficiente o engañosa para determinar las preferencias de los usuarios y porque existen otras investigaciones más profundas que un simple completado de matrices [2].

2.1.3 Aplicaciones

Actualmente los sistemas de recomendación se aplican en varios dominios para resolver diversos problemas [25]. En el dominio de las redes sociales se recomiendan amigos, algunos ejemplos son Facebook, Linkedin y Twitter. En el dominio del comercio

electrónico se recomiendan productos, un ejemplo es Amazon. En el dominio de videos se recomiendan videos, algunos ejemplos son YouTube, Instagram y TikTok. En el dominio de películas se recomiendan películas, algunos ejemplos son Netflix y Amazon Prime. En el dominio de la música se recomiendan canciones, un ejemplo es Spotify. En el dominio del reclutamiento se recomiendan trabajos, algunos ejemplos son Indeed, Careercloud, Job-Hunt, Jibberjobber y Careerbuilder. En el dominio de motores de búsqueda se recomiendan sitios web, por ejemplo, Google. En el dominio del turismo se recomiendan hoteles, restaurantes y lugares, por ejemplo el sistema de [22].

2.1.4 Desafíos

Los principales problemas o desafíos de los sistemas de recomendación se describen en el trabajo de [7], los cuales son los siguientes: (1) La precisión que hace referencia a la eficiencia de los resultados. (2) El arranque en frío es un problema que se presenta con usuarios o ítems nuevos porque el sistema no conoce los intereses de los usuarios y los usuarios no han calificado ningún ítem; en el caso de un nuevo ítem ningún usuario lo ha calificado por lo que no se conoce si es preferido o no por otros usuarios. (3) La escasez de datos es un problema que se presenta en las matrices de calificaciones porque un usuario o ítem no tienen suficiente cantidad de calificaciones lo que reduce el rendimiento de la recomendación. Además, los usuarios no califican todos los productos y si existe un alto número de usuarios e ítems dificultan más este problema. (4) La escalabilidad mide la capacidad de procesamiento de altos volúmenes de datos en los sistemas. Esta capacidad no debe ser afectada por el incremento de información, sin embargo, el incremento de la cantidad de usuarios e ítems representa un mayor número de cálculos, lo que causa un mayor costo de cómputo. (5) La sobre-especialización se presenta en las recomendaciones basadas en el perfil del usuario porque no se descubren nuevos ítems que pertenezcan a otras categorías distintas del perfil del usuario. (6) La diversidad en las recomendaciones para que los ítems provengan de diversas categorías. (7) La novedad en las recomendaciones para que contengan ítems nuevos. (8) La serendipia es más complejo que la novedad porque se busca que las recomendaciones de los ítems sorprendan al usuario, sin embargo, cada usuario tiene gustos diferentes. (9) La privacidad es otro desafío que se presenta porque el sistema requiere conocer los intereses del usuario para poderle realizar recomendaciones. (10) Finalmente, el concepto denominado “Oveja gris” ocurre en el filtrado colaborativo cuando un usuario no pertenece a ningún grupo con gustos similares por lo que el sistema no puede recomendarle ítems.

2.2 Enfoques de recomendación

2.2.1 Filtrado basado en contenido

En este tipo de enfoque se construye un perfil del usuario de forma explícita con formularios donde el usuario expresa sus preferencias o de forma implícita con la extracción de información de los ítems preferidos anteriormente por el usuario. Las recomendaciones se realizan a partir del perfil del usuario. Este enfoque depende del contexto porque los ítems deben tener un conjunto de atributos que los describan, los cuales son llamados metadatos. Los metadatos se especifican manualmente o se obtienen analizando información complementaria como etiquetas, comentarios, descripciones textuales o contenido multimedia, sin embargo, se debe relacionar el formato de los metadatos de cada ítem con el formato del perfil del usuario para estimar el interés que el usuario presenta sobre cada ítem. En la Figura 1, se muestra el procedimiento del filtrado basado en contenido que consiste en lo siguiente: (1) extraer los atributos de los ítems, (2) comparar atributos de los ítems con el perfil del usuario y (3) recomendar los ítems que sean más similares al perfil del usuario [19].

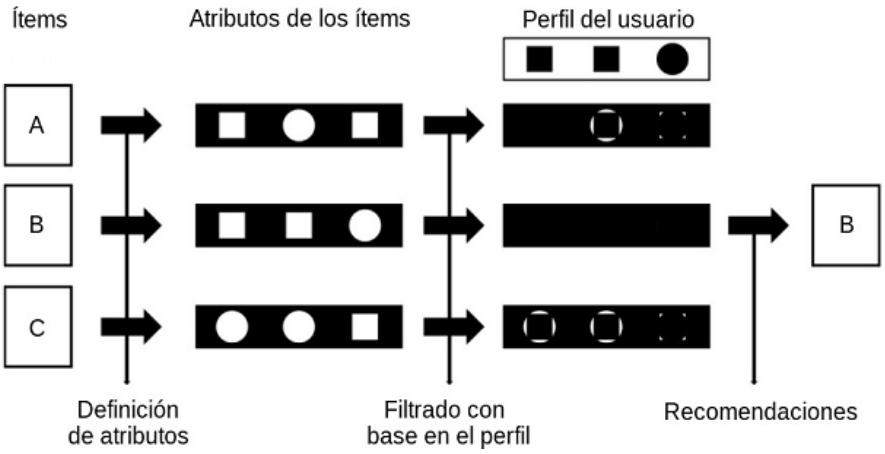

Figura 1. Filtrado basado en contenido [19].

Las ventajas de este enfoque son las buenas estimaciones que presenta en sus resultados por la especificación del perfil del usuario y de los metadatos, y no presenta el problema de arranque en frío. Sin embargo, sus desventajas son la sobre-especialización, la novedad, la diversidad, la complejidad del análisis de datos multimedia que requieren de una descripción textual y el mantenimiento de la información textual de los ítems [14].

2.2.2 Filtrado colaborativo

Son un conjunto de técnicas más populares para el desarrollo de sistemas de recomendación porque aprovecha la opinión de otros usuarios para realizar las recomendaciones. La representación de este enfoque es con una matriz de calificaciones, donde los renglones son los usuarios, las columnas son los ítems y las celdas son las calificaciones de los usuarios hacia los ítems; las calificaciones representan el nivel de preferencia que tiene un usuario hacia un ítem. En la Figura 2, se muestra un ejemplo de la matriz de calificaciones, donde, *u* representa a los usuarios, los ítems son representados por las letras *A, B, C* y *D*, y las calificaciones son evaluadas usando valores que se encuentran en un intervalo de 1 a 5, siendo 1 la peor valoración y 5 la mejor valoración para un ítem.

	A	B	C	D
u_1	3	2	2	1
u_2	3	1	2	5
u_3	∅	5	∅	2
u_4	5	4	2	4
u_5	2	3	4	2

Figura 2. Matriz de calificaciones [19].

Las ventajas de este enfoque en comparación con el del filtrado basado en contenido son que no depende de la descripción de los ítems y no presenta los problemas de diversidad y sobre-especialización. Sin embargo, sus desventajas son el arranque en frío, la escasez de datos, la escalabilidad y la oveja gris [14]. Este enfoque se clasifica a su vez, en basado en memoria o basado en modelo. A continuación, se describen estos dos sub-enfoques.

Basado en memoria

Los algoritmos estiman el interés de un usuario hacia un ítem utilizando la matriz de calificaciones, se emplea una función de agregación para calcular la estimación de los ítems desconocidos para el usuario, es decir, para los ítems que no ha calificado el usuario. Las estimaciones se pueden realizar con la técnica de filtrado colaborativo basado en el usuario o basado en el ítem. La ventaja de este enfoque es que es un proceso simple y eficaz, sin embargo, su desventaja es el costo de cómputo al calcular la similitud entre usuarios o ítems.

Basado en el usuario

La función de agregación se basa en el usuario, dado un usuario se buscan sus usuarios más similares (vecinos cercanos) y se predice la calificación de los ítems no calificados por el usuario con las calificaciones de sus vecinos cercanos. Generalmente, se genera una matriz de similitud de los usuarios, sin embargo, al procesar un alto número de usuarios en una matriz se incrementa el costo de cómputo. El algoritmo que se emplea en este enfoque se llama *k*-vecinos y en la Figura 3, se observa el proceso de este algoritmo que obtiene tres usuarios similares (*U2*, *U5* y *U7*) con respecto del usuario 4, en el ítem 7 dos vecinos lo calificaron con 4 y 5, entonces, el promedio es 4.5 que es la estimación que se predice que valorará el usuario 4 al ítem 7.

Figura 3. Ejemplo del algoritmo *k*-vecinos para el filtrado colaborativo basado en el usuario [19].

Basado en el ítem

La función de agregación se basa en el ítem, dado un ítem no calificado por un usuario, se buscan los ítems más similares (vecinos cercanos) a ese ítem no calificado y se predice su calificación con las calificaciones de los ítems similares evaluados por el usuario. El algoritmo que se emplea en este enfoque se llama *k*-vecinos y en la Figura 4, se observa el proceso de este algoritmo que obtiene tres ítems similares (*I5*, *I9* e *I10*) al ítem 1 del usuario 4, el ítem 9 no se toma en cuenta porque el usuario 4 no lo calificó y sólo se promedian el ítem 10 y 5, el resultado para el ítem 1 y del usuario 4 es una predicción con un valor de 5.

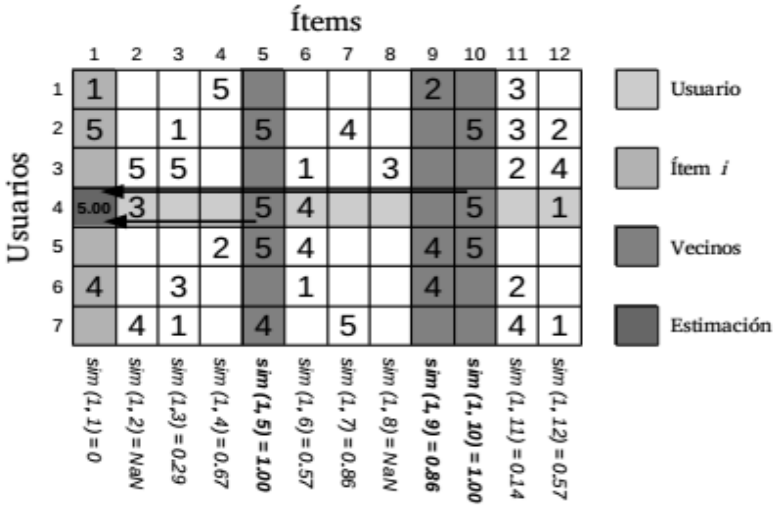

Figura 4. Ejemplo de algoritmo *k*-vecinos basado en el ítem [19].

Las funciones de similitud más usadas para el filtrado colaborativo basado en el usuario o en el ítem son la correlación de Pearson, el coseno, la correlación de rankeo de Spearman y el error cuadrático medio. Las funciones de agregación más usadas son la media y la media ponderada.

Basado en modelo

Este conjunto de técnicas construye un modelo de recomendación con las calificaciones de los usuarios y se formula un problema de datos perdidos o clasificación. Además, se emplean algoritmos de aprendizaje automático como son redes bayesianas, redes neuronales, agrupamiento y factorización de matrices.

2.2.3 Filtrado demográfico

Este enfoque utiliza información del perfil del usuario como la edad, sexo, escuela, intereses y su opinión sobre la calificación de los ítems para encontrar usuarios similares y dividirlos por grupos de edad, sexo o vivienda [7].

2.2.4 Filtrado basado en el conocimiento

Este enfoque se usa para ítems que ocasionalmente se compran porque es imposible recomendar esos ítems que no tienen un historial de compra o mediante la creación de un perfil de usuario. Un ejemplo de estos ítems son los bienes raíces que dificultan la implementación del enfoque colaborativo o basado en contenido. Estas técnicas solicitan detalles específicos y preferencias del usuario para satisfacerlas con las recomendaciones. En el ejemplo de bienes raíces, el sistema solicita información sobre

los requisitos del usuario como son el tipo de inmueble, la localidad, su presupuesto, el número de habitaciones y el número de plantas; con base en esta información el sistema recomienda ítems que cumplan con esas condiciones anteriores. Sin embargo, la principal desventaja de este enfoque es el problema de la novedad y la sobre especialización [28].

2.2.5 Filtrado híbrido

Este enfoque combina varias técnicas para mejorar las recomendaciones y minimizar los problemas particulares de cada enfoque de recomendación. El objetivo de este conjunto de técnicas es preservar las virtudes de cada enfoque de recomendación y que las desventajas de una técnica específica puedan ser minimizadas por las propiedades de los otros enfoques [20] [19] [7].

Ponderado

En esta técnica se combinan las estimaciones de dos o más técnicas de recomendación utilizando una combinación lineal para ser ponderada la relevancia que tendrán las estimaciones hechas por cada enfoque de recomendación que se están combinando al asignar distintos pesos a cada uno de ellos.

Selección

En esta técnica se calcula la estimación de cada una de las técnicas de recomendación a combinar y con base a ciertas condiciones se selecciona el valor para la estimación global.

Mezcla

Esta técnica mezcla las recomendaciones de varios sistemas de recomendación en una lista de recomendaciones global. Además, trabaja sobre las calificaciones generadas a partir de las estimaciones del nivel de interés del usuario sobre los ítems recomendables.

Combinación de características

En este enfoque, se unen múltiples fuentes de conocimiento para construir las características del sistema de recomendación.

Aumento de características

Esta técnica es más flexible y debe abordar la agregación dimensional de los datos de entrenamiento más grandes y agrega un número menor de características a la entrada del enfoque principal.

Cascada

Es una técnica basada en jerarquía, el enfoque más débil no puede revertir las decisiones tomadas por uno más fuerte, sino que simplemente puede refinarlas.

Meta nivel

Esta técnica usa un modelo aprendido por un enfoque como entrada para otro.

2.3 Métricas de evaluación

La evaluación de los algoritmos de recomendación es otro desafío importante para esta área de investigación porque (1) el rendimiento de los algoritmos puede variar con diferentes escalas de los conjuntos de datos, (2) existen varios indicadores de evaluación pero algunos son contradictorios como la precisión y la diversidad; y (3) los diversos indicadores deben medirse de forma diferente. Sin embargo, la evaluación de estos algoritmos es fundamental para analizar las ventajas y desventajas con base en el enfoque de recomendación aplicado. Las métricas de evaluación se clasifican desde la perspectiva del aprendizaje automático, recuperación de información e interacción humano-computadora [29].

2.3.1 Perspectiva del aprendizaje automático

En el estado del arte se identifican algoritmos de aprendizaje automático que se utilizan para predecir las calificaciones de los usuarios hacia los ítems, por tal motivo, la precisión de las predicciones se calcula con las métricas Mean Square Error (MAE), Mean Square Error (MSE) y Root Mean Square Error (RMSE), las fórmulas de estas métricas son las siguientes:

$$MAE = \frac{1}{|Q|} \sum_{(u,i)\in Q} |r_{ui} - \hat{r}_{ui}|$$

$$MSE = \frac{1}{|Q|} \sum_{(u,i)\in Q} (r_{ui} - \hat{r}_{ui})^2$$

$$RMSE = \sqrt{\frac{1}{|Q|} \sum_{(u,i)\in Q} (r_{ui} - \hat{r}_{ui})^2}$$

Donde Q es un conjunto de datos de prueba, r_{ui} representa la calificación real de un ítem calificado por un usuario y $\hat{r}_{ui}$ representa la predicción realizada por el algoritmo

de recomendación. MAE es una métrica simple y no toma en cuenta la dirección del error. MSE tiene una penalización mayor sobre los errores grandes pero el error cuadrado no tiene un intuitivo significado y RMSE es la más utilizada por los sistemas de recomendación.

2.3.1 Perspectiva de recuperación de información

Los sistemas de recomendación están estrechamente relacionados con los sistemas de recuperación de información porque recuperan información sobre datos históricos del usuario. Al usuario no siempre le interesa la precisión numérica de las calificaciones de los ítems sino le interesa conocer sus características de estos. Por tal motivo, son importantes las métricas de soporte de toma de decisiones y ordenamiento. Estas métricas son precisión, recall, F-Measure, mean average precision (MAP), ROC curve, mean reciprocal Rank (MRR), Spearman Rank correlation coefficient (SRCC), normalized discounted cumulative gain (nDCG) y coverage.

Las métricas de precision, recall,y F-Measure se utilizan en las recomendaciones del tipo *top-n*, en donde los sistemas sugieren una lista ordenada de ítems para el usuario. Las fórmulas de estas métricas son las siguientes:

$$Precision = \frac{N_{rs}}{N_s}$$

$$Recall = \frac{N_{rs}}{N_r}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

Donde N_{rs} es el número de ítems recomendados que el usuario prefiere, N_s es el número de ítems recomendados y N_r el número de ítems que el usuario prefiere. La precision describe la proporción de ítems que el usuario prefiere, el recall describe la proporción de los ítems favoritos del usuario que no se pierden y el F-Measure es el compromiso entre precisión y recall. En la aplicación los usuarios se centran más en la precision que en el recall.

La métrica llamada en inglés mean average precisión (MAP) es el promedio de la precision de recomendaciones múltiples, si la precision de los ítems recomendados es alta entonces el MAP también será alto. La fórmula de esta métrica es la siguiente:

$$MAP = \frac{\sum_{q=1}^{Q} Avep(q)}{Q}$$

$$Avep(q) = \frac{\sum_{k=1}^{n} p(k) * rel(k)}{\#relevant\ item}$$

Donde Q es el número de recomendaciones, k es el rango, $rel(k)$ representa la función de relatividad dado el rango k y $p(k)$ representa la precision dada por el rango k.

La métrica ROC curve fue utilizada inicialmente para el análisis de las señales, sin embargo, esta métrica es utilizada en algoritmos de aprendizaje automático y recuperación de información. ROC curve es un gráfico de coordenadas bidimensional, el eje X es la tasa de falsos positivos y el eje Y es la tasa de verdaderos positivos. Además, el rendimiento de diferentes sistemas de recomendación se puede comparar con el área bajo la curva.

La métrica mean reciprocal rank (MRR) es utilizada en los sistemas de recuperación de información para clasificar los resultados más relevantes recuperados, es decir, en los algoritmos de recomendación se puede medir si los ítems favoritos del usuario están al frente de una lista. La fórmula de esta métrica es la siguiente:

$$MRR = \frac{\sum_{q=1}^{Q} \frac{1}{rank_i}}{Q}$$

Donde Q es el número de recomendaciones, $rank_i$ son los ítems favoritos del usuario en la lista ordenada de recomendaciones. Un valor alto para MRR representa un mejor rendimiento del sistema de recomendación.

La métrica Spearman rank correlation coefficient (SRCC) es usada para calcular el coeficiente de correlación de Pearson entre el orden de los ítems recomendados y el ground truth. La fórmula de esta métrica es la siguiente:

$$SRCC = \frac{\sum_i (r_1(i) - \mu_1)(r_2(i) - \mu_2)}{\sqrt{\sum_i (r_1(i) - \mu_1)^2} \sqrt{\sum_i (r_2(i) - \mu_2)^2}}$$

Donde $r_1(i)$ y $r_2(i)$ son la clasificación de los ítems recomendados y el ground truth, y μ es el promedio de las calificaciones de los ítems. Si la calificación de los ítems recomendados es la misma que la calificación real, entonces el valor de SRCC es 1.

La métrica normalized discounted cumulative gain (nDCG) es ampliamente utilizada en la evaluación de los sistemas de recomendación en comparación de SRCC, porque SRCC no considera el orden de las recomendaciones. La fórmula de esta métrica es la siguiente:

$$nDCG = \frac{DCG(r)}{DCG(r_{perfect})}$$

$$DCG(r) = \sum disc(r(i))u(i)$$

Donde $disc(r(i))$ es una función descontinua basada en las calificaciones de los ítems, $u(i)$ es la utilidad de los ítems en la lista de recomendación y $DCG(r_{perfect})$ representa una perfecta clasificación de la ganancia acumulada descontada.

La métrica coverage es la proporción de ítems recomendados sobre el total de ítems. La fórmula de esta métrica es la siguiente:

$$Coverage = \frac{U_{u \epsilon U} I(u)}{I}$$

Donde $I(u)$ es el número de ítems recomendados para un usuario e I representa el número total de ítems. El coverage es una métrica importante porque puede describir la capacidad de extraer el rango de ítems recomendados.

Capítulo 3: Revisión sistemática de la literatura en los enfoques híbridos de recomendación

La revisión de la literatura implica detectar, consultar y obtener las referencias que sean útiles para el propósito del estudio. De estas referencias se extrae y recopila la información relevante para el problema de investigación [10]. En este capítulo se presenta una revisión sistemática de la literatura que es un método sistemático, explicito y reproducible para identificar, evaluar y sintetizar investigaciones completas y registradas producidas por investigadores, académicos y profesionales [30]. A diferencia de otros tipos de revisiones de la literatura enfocadas al desarrollo de un marco teórico de una investigación como parte de un artículo o un capítulo de tesis, una revisión sistemática de la literatura es una investigación más amplia como un artículo completo de investigación que define altos criterios sistematizados de inclusión, exclusión y calidad para la revisión de la literatura para desarrollar un estado del arte sistematizado y reproducible.

La revisión sistemática de la literatura para enfoques híbridos de recomendación se construye con la guía se Okoli y Schabram [30] que es dirigida a sistemas de información un área estrechamente relacionada a sistemas de recomendación a diferencia de otros trabajos como [31], [32] y [33] que utilizan guías enfocadas a la ingeniería de software y que son guías menos recientes. El principal trabajo relacionado es el Çano y Morisio [31] porque es una revisión sistemática de la literatura específicamente para sistemas híbridos de recomendación que desarrolla un estado del arte de la década del 2005 al 2015, por tal motivo, en este estudio se desarrolla un estado del arte que aborda el vacío de conocimiento del último lustro del 2016 al 2020. La guía de Okoli y Schabram [30] está formada por ocho pasos secuenciales que son el propósito de la revisión de la literatura, el protocolo y la capacitación, la búsqueda de la literatura, el proceso de selección, la evaluación de calidad, la extracción de datos, el análisis de los resultados y síntesis de la revisión. A continuación, se desarrollan estos ocho pasos secuenciales.

3.1 Propósito

El propósito general de esta revisión sistemática de la literatura es analizar el progreso de los enfoques híbridos de recomendación e identificar sus áreas de oportunidad para futuras investigaciones. Este trabajo se centra en el análisis de las tendencias actuales sobre desafíos, metodologías como técnicas, algoritmos, métodos y experimentos, conjuntos de datos, formas de evaluación y trabajos futuros para enfoques híbridos de recomendación.

3.2 Protocolo y capacitación

Las preguntas de investigación para esta revisión sistemática de la literatura para enfoques híbridos de recomendación son las siguientes:

1. ¿Cuáles son los estudios recientes y relevantes sobre enfoques híbridos de recomendación?
2. ¿Cuáles problemas abordan los enfoques híbridos de recomendación?
3. ¿Qué soluciones experimentales se han generado con enfoques híbridos de recomendación?
4. ¿Qué técnicas y algoritmos se utilizan para enfoques híbridos de recomendación?
5. ¿Qué formas y métricas de evaluación se utilizan para enfoques híbridos de recomendación?
6. ¿Qué conjuntos de datos se usan para enfoques híbridos de recomendación?
7. ¿Cuáles son los dominios de aplicación de los enfoques híbridos de recomendación?
8. ¿Cuáles son los trabajos futuros y las áreas de oportunidad para los enfoques híbridos de recomendación?

3.3 Búsqueda de la literatura

Las fuentes de información que se seleccionan son cinco principales bases de datos científicas que se han utilizado en otros trabajos relacionados publicados en revistas indizadas en Journal Citation Report (JCR) [31] [32]. Estas bases de datos son: *ACM Digital Library, IEEE Xplore, Scopus, Springer Link y Web of Science.*

En la construcción de la consulta para las bases de datos previamente seleccionadas se usan seis palabras clave acompañadas de un conjunto de sinónimos por cada palabra clave. Las primeras tres palabras clave se obtuvieron del trabajo [31] y las otras tres palabras clave se proponen en esta investigación para abarcar temas de desafíos, conjuntos de datos y evaluaciones sobre enfoques híbridos de recomendación. En la

Tabla 2, se describen las palabras clave con sus sinónimos que se definieron como términos para la consulta en las bases de datos.

Tabla 2. Palabras clave y sinónimos.

No	Palabra clave	Sinónimos
1	Hybrid	Hybridization, Mixed
2	Recommender	Recommendation
3	System	Systems, Software, Technique, Techniques, Technology, Approach, Engine
4	Challenge	Challenges, Problem, Problems, Issue, Trouble
5	Dataset	Corpus
6	Evaluation	Assessment, Metrics

La consulta que se define con los términos previamente establecidos se forma con operadores booleanos AND y OR, se unen dos subconsultas generales con el operador OR, la primera subconsulta representa la búsqueda general sobre sistemas de recomendación híbridos y la segunda subconsulta representa la búsqueda especifica de los desafíos, conjuntos de datos y evaluaciones sobre enfoques híbridos de recomendación. La consulta que se genera se muestra en la Tabla 3.

Tabla 3. Consulta para bases de datos.

Consulta
((Hybrid OR Hybridization OR Mixed) AND (Recommender OR Recommendation) AND (System OR Systems OR Software OR Technique, OR Techniques OR Technology OR Approach OR Engine)) OR ((Hybrid OR Hybridization OR Mixed) AND (Recommender OR Recommendation) AND (System OR Systems OR Software OR Technique, OR Techniques OR Technology OR Approach OR Engine) AND (Challenge OR Challenges OR Problem OR Problems OR Issue OR Trouble OR Dataset OR Corpus OR Evaluation OR Assessment OR Metrics))

La búsqueda de la literatura se realiza con las cinco bases de datos previamente seleccionadas y con la ejecución de la consulta previamente definida, los resultados obtenidos en general son de 383,937 publicaciones recuperadas, de las cuales, en Springer Link se recuperaron 319,124 publicaciones, en ACM Digital Library se recuperaron 37,349 publicaciones, en Scopus se recuperaron 12,921 publicaciones, en Web of Science se recuperaron 12,717 publicaciones y en IEEE Xplore se recuperaron 1,826 publicaciones. La base de datos que obtuvo un mayor número de publicaciones recuperadas es Springer Link y la base de datos que obtuvo un menor número de publicaciones recuperadas es IEEE Xplore, sin embargo, el proceso de selección y los

criterios de calidad determinarán que publicaciones se seleccionarán para generar el estado del arte de esta investigación.

3.4 Proceso de selección

En el proceso de selección se definen los criterios de inclusión y exclusión de material para seleccionar las publicaciones que se revisaron por primera vez de las 382,937 publicaciones recuperadas en la búsqueda de la literatura.

Los criterios que se definen para incluir el material de las publicaciones recuperadas de las bases de datos científicas se describen en la Tabla 4.

Tabla 4. Criterios de inclusión.

Materia incluido
El titulo y el resumen del articulo deben centrarse sobre enfoques híbridos de recomendación.
Las publicaciones deben ser artículos de revistas o conferencias.
Los artículos deben estar publicados entre el año 2016 al 2020 para cubrir un vacío de información debido al principal y más reciente trabajo relacionado de Çano y Morisio [31] que incluye artículos publicados entre el año 2005 al 2015.
Artículos redactados en el idioma inglés.

Los criterios que se definen para excluir el material de las publicaciones recuperadas de las bases de datos científicas se describen en la Tabla 5.

Tabla 5. Criterios de exclusión.

Material excluido
El titulo y el resumen del artículo que se centren en temas diferentes a los enfoques híbridos de recomendación.
Las publicaciones que no son artículos de revistas o conferencias como la literatura gris.
Los artículos publicados antes del año 2016 y después del año 2020.
Artículos redactados en otros idiomas diferentes al idioma inglés.

En esta primera revisión de la literatura se identificaron 70 artículos que cumplen con los criterios de inclusión y exclusión para el proceso de selección previamente definido,

de los cuales, 15 artículos se seleccionaron de ACM Digital Library, 25 artículos se seleccionaron de Web of Science, 19 artículos se seleccionaron de Springer Link, 6 artículos se seleccionaron de Scopus y 5 artículos se seleccionaron de IEEE Xplore.

3.5 Evaluación de calidad

La calidad de las publicaciones se evalúa con la estrategia de Çano y Morisio [31] que consiste en un cuestionario ponderado, en el cual las preguntas más importantes tienen un peso de 1.5, las preguntas con una importancia intermedia tienen un peso de 1 y las preguntas con menos importancia tienen un peso de 0.5. El valor de las respuestas es asignado por el revisor del artículo, si la respuesta es afirmativa (Si) el valor es 1, si la respuesta es parcialmente el valor es de 0.5 y si la respuesta es negativa (No) el valor es de 0. En la Tabla 6, se muestra los criterios de calidad con base en 7 preguntas y sus valores para las respuestas ponderadas.

Tabla 6. Criterios de calidad.

#	Pregunta	Valor	Peso
1	¿El estudio presenta de forma clara el problema que aborda?	0;0.5;1	1
2	¿El estudio desarrolla de forma clara una solución experimental?	0;0.5;1	1.5
3	¿El estudio describe las técnicas y algoritmos de recomendación que utiliza?	0;0.5;1	1
4	¿El estudio presenta una evaluación de sus experimentos con métricas de sistemas de recomendación?	0;0.5;1	1.5
5	¿El estudio describe el conjunto de datos que utilizó?	0;0.5;1	1
6	¿El estudio presenta el dominio de aplicación?	0;0.5;1	0.5
7	¿El estudio presenta trabajos futuros con enfoques híbridos de recomendación?	0;0.5;1	0.5

La fórmula para evaluar cada artículo se presenta a continuación.

$$\sum_{i=1}^{7} p_i * v_i \,/7$$

Esta fórmula realiza el producto del peso de la pregunta con la valoración del revisor y se calcula el promedio entre las 7 preguntas. El umbral de calidad para aceptar los artículos debe ser mayor de 0.81. Una segunda revisión de la literatura es realizada con

base en los criterios de calidad para seleccionar 36 artículos de los 70 artículos seleccionados de la primera revisión de la literatura. La segunda revisión de la literatura consiste en la revisión a detalle de los 36 artículos por medio de lecturas completas, se seleccionan 12 artículos de ACM Digital Library, 12 artículos de Web of Science, 5 artículos de Springer Link, 4 artículos de IEEE Xplore y 3 artículos de Scopus. En la Tabla 7, se presenta la cantidad de artículos recuperados en las bases de datos, la cantidad de artículos seleccionados en la primera revisión y en la segunda revisión de la literatura.

Tabla 7. Artículos recuperados y revisados.

Base de datos	Recuperación	1a. Revisión	2a. Revisión
ACM Digital Library	37,349	15	12
Web of Science	12,717	25	12
Springer Link	319,124	19	5
IEEE Xplore	1,826	5	4
Scopus	12,921	6	3
TOTAL	383,937	70	36

3.6 Extracción de datos

La extracción de datos se realiza sobre los 36 artículos seleccionados por la segunda revisión de la literatura, se genera un formulario para la extracción de 18 datos que describen en la Tabla 8, estos datos sirven para contestar las preguntas de investigación de este trabajo, por tal motivo, se relaciona cada extracción de datos con una pregunta de investigación en la cuarta columna.

Tabla 8. Formulario para la extracción de datos.

#	Extracción de datos	Descripción	Respuesta a p. de inv.
1	id	Número de identificación.	-
2	Titulo	-	R1
3	Autores	-	-
4	Año	-	R1
5	Nombre	Nombre de revista o conferencia.	-
6	Volumen	Volumen de la revista.	-

7	Número	Número de la revista.	-
8	Lugar	Lugar de publicación.	-
9	Fuente de datos	Nombre de base de datos.	-
10	Problema	Problema de investigación.	R2
11	Experimento	Descripción experimental.	R3
12	Métodos	Algoritmos y métodos aplicados.	R4
13	Técnicas	Técnicas individuales de recomendación.	R4
14	Enfoque híbrido	Enfoque híbrido aplicado.	R4
15	Evaluación	Formas y métricas de evaluación.	R5
16	Conjunto de datos	Dataset empleado.	R6
17	Dominio	Dominio de aplicación.	R7
18	Trabajo futuro	Sugerencias de trabajos futuros.	R8

3.7 Análisis de los resultados y síntesis de la revisión

La distribución de los 16 artículos de revistas y los 20 artículos de congresos que son incluidos en esta revisión sistemática de la literatura para enfoques híbridos de recomendación se muestra en la Figura 5, en donde se clasificó en orden cronológico la cantidad y el tipo de publicaciones analizadas por cada año para demostrar que los artículos incluidos en esta revisión sistemática son recientes y actuales. Además, se muestra una mínima diferencia entre la cantidad de artículos de revistas y congresos debido a que la mayoría de los artículos de revistas son más extensos que los artículos de congresos.

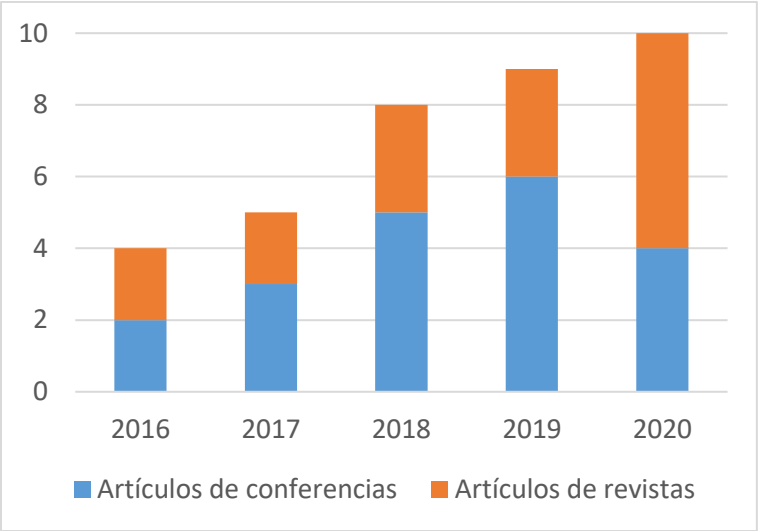

Figura 5. Distribución de artículos incluidos por cada año.

Los artículos de revistas presentan mejores resultados en la evaluación de calidad que los artículos de congresos, en la Figura 6 se muestra la calidad promedio para ambos tipos de publicaciones, en el cual la calidad promedio para los artículos de revistas es de 0.97 y la calidad promedio para los artículos de congresos es de 0.88, sin embargo, ambos cumplen con los criterios de calidad establecidos y este análisis sirve para clasificar los artículos por su nivel de calidad para desarrollar un estado del arte más sólido.

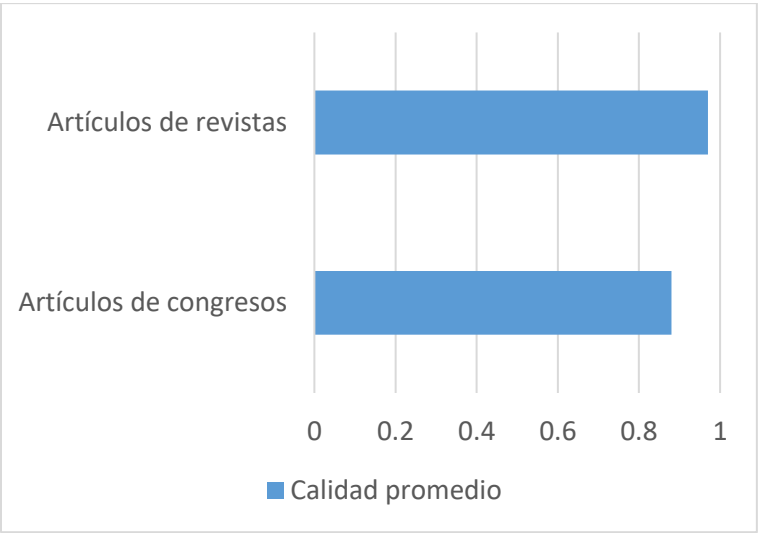

Figura 6. Calidad promedio de artículos incluidos de revistas y congresos.

3.7.1 Problemas y desafíos

La distribución de los problemas que se abordan con enfoques híbridos de recomendación en los 36 estudios analizados se muestra en la Figura 7, en la cual se demuestra que los seis desafíos más abordados en el estado del arte son el arranque en frío, la exactitud, la escasez de datos, la integración de fuentes de información (F.I), la escalabilidad y otros problemas como la falta de personalización, las recomendaciones basadas en aspectos, las recomendaciones basadas en localización, la falta de novedad, la explicación de recomendaciones, la transparencia e interoperabilidad, las recomendaciones basadas en sesiones y el paralelismo en algoritmos de recomendación. En total se identificaron 13 problemas de investigación sobre enfoques híbridos de recomendación y se clasificaron en las 6 principales tendencias recientes sobre estos desafíos para enfoques híbridos de recomendación.

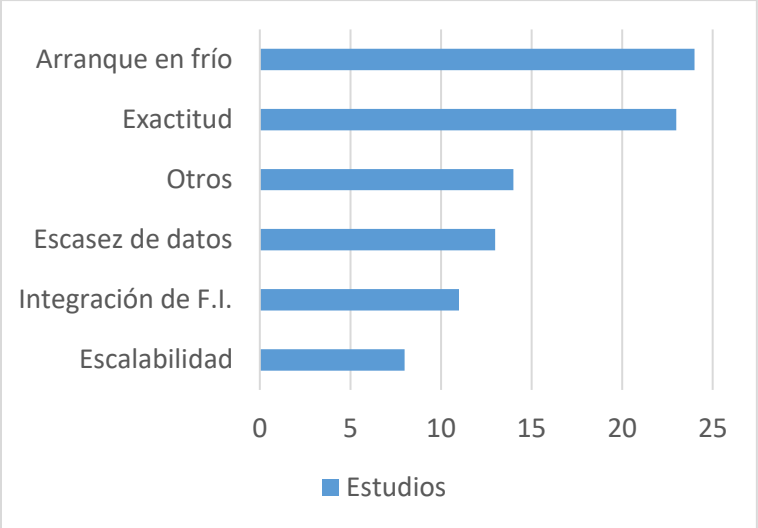

Figura 7. Distribución de los problemas actuales sobre enfoques híbridos de recomendación.

La distribución de las soluciones actuales para los problemas identificados sobre enfoques híbridos de recomendación se presenta en la Tabla 9, en la cual se muestran los desafíos que se abordan con un específico estudio, es importante señalar que solo 3 estudios abordan un único problema de investigación, mientras que la mayoría de los estudios abordan dos o más problemas de investigación gracias a la combinación de algoritmos y enfoques de recomendación, por ejemplo en los estudios [34] y [35] se abordan hasta 5 problemas de recomendación.

Tabla 9. Distribución de las soluciones actuales para los problemas sobre enfoques híbridos de recomendación.

ID	Estudio	Arranque en frío	Exactitud	Otros	Escasez de datos	Integración de fuentes de información	Escalabilidad
1	[36]		x	x		x	
2	[37]	x					x
3	[38]	x	x				
4	[39]	x	x				
5	[40]	x	x				
6	[41]	x	x	x	x		
7	[42]	x			x	x	
8	[43]		x	x		x	
9	[44]	x			x		
10	[45]		x				
11	[46]	x	x		x		
12	[24]	x	x				
13	[47]		x	x		x	
14	[48]	x					
15	[49]	x					
16	[50]	x			x	x	
17	[51]	x	x	x			
18	[52]		x				x
19	[53]		x	x			

20	[54]	x			x		
21	[55]		x	x			x
22	[56]		x			x	x
23	[57]	x				x	
24	[58]	x	x			x	x
25	[59]	x			x		
26	[60]		x	x			x
27	[61]	x		x			
28	[62]	x	x				
29	[63]	x	x				
30	[64]	x			x		
31	[65]		x	x	x		
32	[66]	x		x	x		
33	[67]		x	x		x	
34	[34]	x	x	x	x	x	
35	[68]				x	x	x
36	[35]	x	x	x	x		x

Arranque en frío

El arranque en frío es un problema tradicional en sistemas de recomendación y es la tendencia principal del estado del arte. Este desafío sucede cuando un usuario o ítem nuevo no tiene un registro del historial de calificaciones en ingles ratings, por tal motivo, predecir las calificaciones de un usuario objetivo hacia un ítem es una tarea difícil sin contar con datos personalizados como referencia [37]. En [34] se aborda este problema con un novedoso sistema de recomendación llamado Capsmf que utiliza un modelo de análisis de texto basado en aprendizaje profundo. En [51] se propone un reordenamiento personalizado para recomendaciones de artículos de investigación usando el contenido del artículo y el comportamiento del usuario. En [40] se propone

un sistema de recomendación híbrido basado en la técnica de expansión de perfil del usuario para aliviar el arranque en frío. En total 24 estudios abordan este problema.

Exactitud

La segunda tendencia es el problema de exactitud que representa la calidad y la eficiencia de las recomendaciones, una alta exactitud en los resultados de las sugerencias significa una mejor calidad de las recomendaciones y representa pocos errores en un sistema [32]. En [36] se aborda este problema con la generación y comprensión de explicaciones personalizadas en sistemas de recomendación híbridos. En [46] se propone un sistema de recomendación híbrido para recomendaciones secuenciales combinando modelos de similitud con cadenas de Markov. En [47] se genera un análisis comparativo de sistemas de recomendación basado en la extracción de opiniones de aspectos sobre reseñas de los usuarios. En total 23 estudios abordan este problema.

Escasez de datos

El tercer problema más abordado en el estado del arte es la escasez de datos que es un problema similar al arranque en frío, este problema sucede porque los usuarios generalmente califican un número bastante limitado de ítems y crece este problema con un catálogo de ítems grande, por tal motivo se presenta una matriz de calificaciones escaza con datos insuficientes para identificar usuarios o ítems similares lo que causa un negativo impacto en la calidad de las recomendaciones [31]. En [66] se aborda este problema con un sistema de recomendación híbrido basado en localización para movilidad y planeación de viajes. En [54] se propone un algoritmo eficiente para sistemas de recomendación utilizando técnicas de Kernel Mapping. En [41] se presenta un novedoso modelo para un sistema de recomendación de hospitales que usa el filtrado híbrido y las técnicas de Big data. En total 13 estudios abordan este problema.

Integración de fuentes de información

Una tendencia importante es el problema de la integración de diversas fuentes de información para enriquecer la información de entrada para un sistema de recomendación y así poder mejorar su rendimiento [36]. En [67] se aborda este problema con un modelo de recomendación explicable neuronal con base en atributos y reseñas textuales. En [57] se propone la continuación automática de una lista de reproducción usando un sistema de recomendación híbrido que combina características de texto y audio. En [42] se presenta un sistema de recomendación para aplicaciones multiplataforma con el modelado de calificaciones y texto. En total 11 estudios abordan este problema.

Escalabilidad

La escalabilidad es una característica difícil de alcanzar porque se relaciona con el número de ítems y usuarios que puede soportar un sistema de recomendación, un sistema diseñado para sugerir pocas recomendaciones a unos cientos de usuarios probablemente no recomendará cientos de ítems a millones de usuarios a menos que este diseñado para ser altamente escalable [31]. En [56] se aborda este problema con la combinación de aspectos de algoritmos genéticos con la hibridación ponderada. En [68] se propone un sistema de recomendación basado en el filtrado colaborativo usando ontologías y técnicas de reducción de dimensionalidad. En [52] se presenta una mejora de la estabilidad de la recomendación del filtrado colaborativo a través de la agrupación bioinspirada con métodos de conjuntos. En total 8 estudios abordan este problema.

Otros problemas

La falta de personalización en las recomendaciones limita la eficiencia de los sistemas porque los usuarios tienen diferentes y diversos gustos, este problema se aborda en [65] con una recomendación híbrida personalizada para grupos de usuarios en un dominio de multimedia, en total 6 estudios abordan este problema. La recomendación basada en aspectos utiliza las reseñas textuales como una rica fuente de información para las preferencias de los usuarios porque contienen las opiniones de los usuarios hacia los aspectos del ítems, en [47] se utilizan técnicas de procesamiento del lenguaje natural (PLN), minería de datos y el análisis de sentimientos para la extracción de aspectos y la recomendación, en total 3 estudios abordan este problema. Las recomendaciones basadas en localización que utilizan información de redes sociales para sugerir ubicaciones de interés se aborda en [66] con un sistema recomendación híbrido basado en localización para movilidad y planeación de viajes en un dominio turístico, en total 2 estudios alivian este problema. Los desafíos abordados por un solo estudio son la falta de novedad, la explicación de recomendaciones, la transparencia e interoperabilidad, las recomendaciones basadas en sesiones y el paralelismo en algoritmos de recomendación. En total 14 estudios abordaron diversos problemas de enfoques híbridos de recomendación.

3.7.2 Técnicas de recomendación

La distribución de las técnicas de recomendación utilizadas para el desarrollo de enfoques híbridos de recomendación se presentan en la Figura 8, en la cual se demuestra que las 7 técnicas de recomendación más usadas son el filtrado colaborativo, el filtrado basado en contenido, el filtrado basado en conocimiento, los sistemas de recomendación basados en secuencia, el filtrado demográfico, los sistemas de recomendación basados en contexto y otras técnicas como son las basadas en utilidad,

en localización, en popularidad, en redes sociales, en el comportamiento, en sesiones, en aspectos, en confianza, en bots y las recomendaciones multiplataforma. En total se identificaron 16 técnicas de recomendación y se clasificaron en 7 principales tendencias sobre estas técnicas para enfoques híbridos de recomendación.

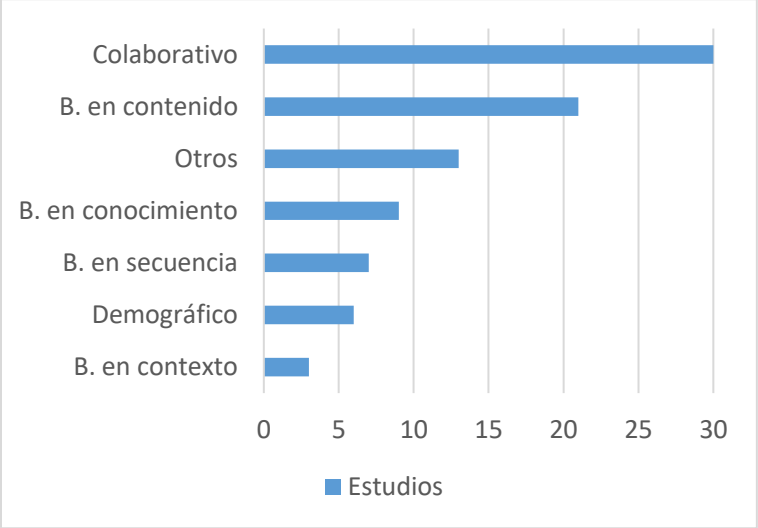

Figura 8. Distribución de las técnicas de recomendación individuales utilizadas sobre enfoques híbridos de recomendación.

Filtrado colaborativo

La técnica más usada con enfoques híbridos de recomendación es el filtrado colaborativo que realiza las recomendaciones para el usuario objetivo con base en opiniones de otros usuarios con gustos similares llamados vecinos cercanos, estas opiniones se representan en una matriz de calificaciones que tiene como renglones a los usuarios, como columnas a los ítems y como valor la calificación. En [40] y [68] se clasifica este enfoque en basado en memoria y basado en modelo, las técnicas basadas en memoria buscan un conjunto de usuarios o ítems similares (vecinos cercanos) con una función de similitud como el coeficiente de la correlación de Pearson o el coseno y se predicen las calificaciones de los ítems que no ha visto el usuario activo con el conjunto de sus vecinos cercanos. Las técnicas basadas en modelo desarrollan un modelo basado en un subconjunto de datos llamado conjunto de entrenamiento y usan modelos para predecir los ítems que se encuentran en los datos restantes llamados conjunto de prueba. En [42] clasifica el filtrado colaborativo en métodos basados en vecinos cercanos y en métodos de factor latente, los métodos de factor latente se centran en ajustar la matriz de calificaciones usando aproximaciones de bajo rango e identificando nuevas asociaciones entre usuarios e ítems. En total 30 estudios utilizan el filtrado colaborativo.

Filtrado basado en contenido

La segunda técnica más utilizada con enfoques híbridos de recomendación es el filtrado basado en contenido que recomienda ítems similares [65], se genera un perfil del usuario con los atributos y las descripciones de los ítems previamente evaluados por el usuario y se buscan los ítem más similares al perfil del usuario [24], esta estrategia se usa generalmente para recursos basados en textos mediante soluciones propuestas en los campos del procesamiento del lenguaje natural (PLN), la recuperación de información y la minería de datos. En total 21 estudios utilizan el filtrado basado en contenido.

Filtrado basado en conocimiento

La tercera técnica más utilizada es el filtrado basado en conocimiento que depende de patrones y reglas extraídas de un conocimiento profundo de los ítems interesantes para el usuario, esto se realiza con una comprensión profunda del mercado y el contexto de aplicación o a través de diálogos o preguntas en algunas aplicaciones y con base en estas preferencias se construye un árbol de discriminación de los atributos de los ítems [49]. Este enfoque utiliza información semántica, por ejemplo, la información semántica de un ítem incluye sus atributos, relaciones entre ítems y la relación entre meta información e ítems [68]. En total 9 estudios utilizan el filtrado basado en conocimiento.

Basado en secuencia

Estas técnicas modelan las dependencias secuenciales sobre las interacciones entre usuarios e ítems como vistas, clics, agregar al carrito, compra y otras interacciones en una secuencia. A diferencia del filtrado colaborativo y basado en contenido que modelan las interacciones entre usuarios e ítems de forma estática, las técnicas basadas en secuencia tratan las interacciones como una secuencia dinámica y analizan las dependencias secuenciales para obtener la preferencia actual y reciente de un usuario para una recomendación más precisa. Estas técnicas se clasifican en modelos de secuencia tradicionales, en modelos de representación latente y en modelos de redes neuronales profundas [69]. En total 7 estudios utilizan recomendaciones basadas en secuencia.

Demográfico

Esta técnica utiliza un perfil demográfico del usuario con información específica como la edad, el generó, el país y otros datos del usuario, se agrupan los usuarios similares y se realizan las recomendaciones con base en esos grupos[61]. En total 6 estudios utilizan el enfoque demográfico.

Sensible al contexto

La información de contexto como el tiempo y los aspectos de comportamiento del usuario como vistas, clics, compras, etc. se han utilizado para enriquecer otras técnicas de recomendación convencionales porque estas características de contexto aumentan el rendimiento de los sistemas de recomendación [35]. Las técnicas sensibles al contexto se clasifican en pre-filtrado, post-filtrado y modelado contextual [42]. En el pre-filtrado el contexto controla la selección de datos, en el pos-filtrado el contexto es usado para filtrar recomendaciones después de una técnica tradicional y en el modelado contextual el contexto es integrado directamente en un modelo. En total 3 estudios utilizan técnicas sensibles al contexto.

Otras técnicas

En total 13 estudios utilizan otras técnicas de recomendación como las basadas en utilidad, en localización, en popularidad, en redes sociales, en el comportamiento, en sesiones, en aspectos, en confianza, en bots y las recomendaciones multiplataforma.

3.7.3 Enfoques híbridos de recomendación

La distribución de los enfoques híbridos de recomendación aplicados en los estudios incluidos en esta revisión sistemática se presenta en la Figura 9, en la cual se demuestra que las técnicas de hibridación más utilizadas son la ponderación, la mezcla, la selección, la combinación de características, el aumento de características, la cascada y otros enfoques nuevos generados en el resto de los estudios con base en la combinación de técnicas, métodos y algoritmos multidisciplinarios. Se clasificaron estos 7 enfoques híbridos con base en la tendencia de los estudios que los utilizan.

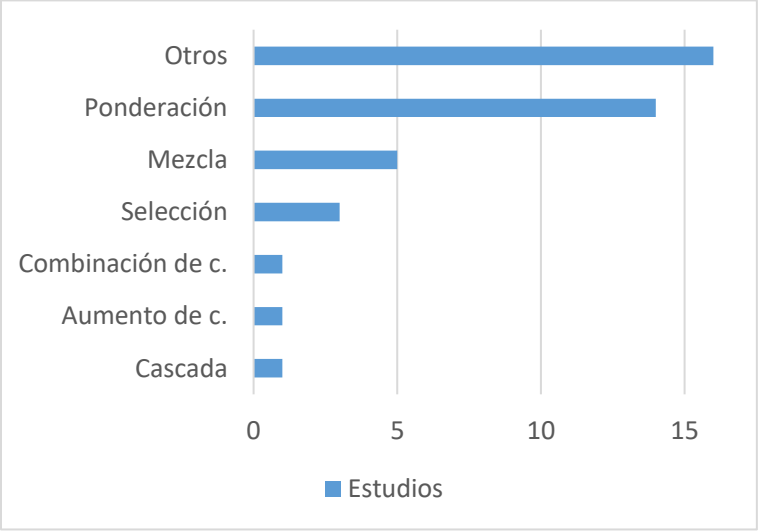

Figura 9. Distribución de los enfoques híbridos de recomendación aplicados en los estudios incluidos.

Ponderación

El enfoque híbrido más utilizado es la ponderación que combina los pesos de varias técnicas de recomendación para producir una única recomendación usando una combinación lineal o un esquema de votación [38]. En total 14 estudios utilizan el enfoque de ponderación.

Mezcla

El segundo enfoque híbrido más utilizado es el de mezcla que utiliza un algoritmo básico para combinar las listas de recomendaciones de varios algoritmos en una sola lista de recomendación [39], el desafío es la presentación y la dificultad de incluir calificaciones en listas combinadas, sin embargo, la mayoría de las combinaciones se basan en la predicción de calificaciones [54]. En total 5 estudios utilizan el enfoque por mezcla.

Selección

El tercer enfoque híbrido más utilizado es el de selección que permite seleccionar una técnica de recomendación entre varias técnicas dependiendo de la situación [39]. En total 3 estudios utilizan el enfoque por selección.

Combinación de características

El enfoque híbrido de combinación de características permite datos de diferentes técnicas para ser combinados y procesados hacía otra técnica de recomendación [39]. Un estudio utilizo el enfoque de combinación de características.

Aumento de características

El enfoque híbrido de aumento de características usa el resultado de una técnica de recomendación como datos de entrada para otra técnica [39]. Un estudio utilizó el enfoque de aumento de características.

Cascada

En enfoque híbrido de cascada jerarquiza las técnicas de recomendación para refinar los resultados de cada técnica, la primera técnica es la principal porque otros enfoques solo refinaran sus resultados [39]. Un estudio utilizó el enfoque de cascada.

Otros enfoques híbridos

En [50] se propone un sistema de recomendación híbrido para redes sociales basado en una estructura topológica hipergráfica que combina el filtrado colaborativo con el basado en contenido utilizando un modelo de factorización de matrices híbrido. En [51] se genera un modelo de recalificación híbrido que combina las recomendaciones basadas en contenido, basadas en comportamiento y basadas en secuencia. En [67] se desarrolla un modelo de recomendación neuronal explicable basado en atributos y reseñas textuales. En [34] se integra un modelo de análisis de texto de red neuronal profunda con la factorización de matrices probabilística. En total 16 estudios presentan otros enfoques híbridos de recomendación, de los cuales los más recientes y con mejor evaluación de calidad usan el aprendizaje profundo para generar enfoques híbridos de recomendación.

3.7.4 Métodos y algoritmos

La distribución de los métodos y algoritmos empleados para el desarrollo de enfoques híbridos de recomendación se presenta en la Figura 10, en la cual se demuestra que los métodos y algoritmos más utilizados son la agrupación, la similitud del coseno, la factorización de matrices, el procesamiento del lenguaje natural (PLN), los algoritmos de optimización, los algoritmos de vecinos cercanos, las redes neuronales, la recuperación de información, la distancia euclidiana y otros métodos o algoritmos como los sistemas basados en reglas, la expansión del perfil del usuario, los algoritmos paralelos, la retroalimentación de las calificaciones, los algoritmos semánticos, la similitud demográfica, la correlación de Pearson, la formula del semiverseno, las cadenas de Markov, las reglas de asociación, las ontologías, los métodos de conjunto, los métodos de Kernel Mapping, la lógica difusa, los agentes inteligentes, árboles de decisión y el proceso analítico jerárquico. En total se identificaron 26 métodos y se clasificaron en 10 principales tendencias sobre estos métodos y algoritmos para enfoques híbridos de recomendación.

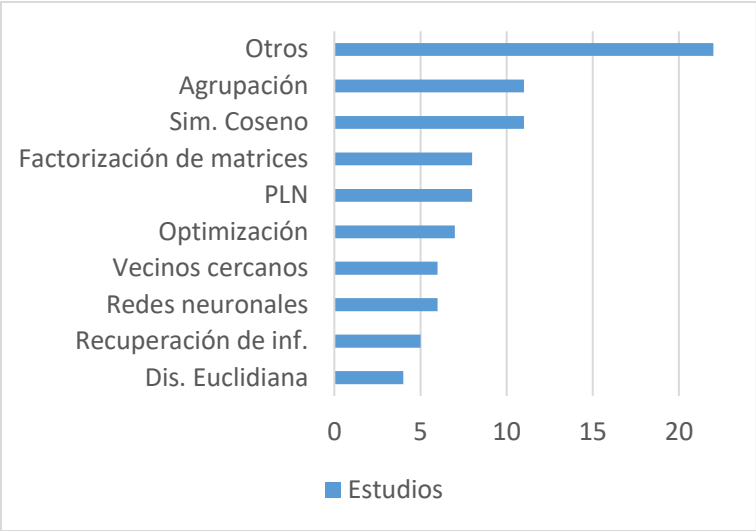

Figura 10. Distribución de los métodos y algoritmos utilizados para los enfoques híbridos de recomendación.

Los algoritmos de agrupación son el método más utilizado en los estudios y generalmente agrupan usuarios similares o ítems similares con el aprendizaje no supervisado, en [50] se utiliza el algoritmo k-modos para agrupar conjuntos de datos de usuarios e ítems basados en información contextual, en [68] se agrupan calificaciones de usuarios sobre películas usando un algoritmo de maximización de expectativas y en [52] se propone un sistema de recomendación de filtrado colaborativo basado en el ensamble de algoritmos de agrupamiento bioinspirado como son K-PSO, FCM-PSO y K-MWO. En total 11 estudios usan algoritmos de agrupación.

La similitud del coseno es el segundo método más utilizado y sirve para calcular la similitud entre usuarios o ítems. En [51] se calcula la similitud del contenido de artículos de investigación utilizando un espacio de palabras y entidades, se usa la similitud del coseno de los vectores que representan a cada artículo, en [47] se emplea la similitud del coseno de diferentes formas para el filtrado basado en contenido dependiendo de la extracción de la opinión de los aspectos del ítem y del perfil del usuario y en [38] se usa el coseno ajustado para buscar ítems similares con el filtrado colaborativo basado en el ítem. En total 11 estudios usan la similitud del coseno.

La factorización de matrices es el tercer método más utilizado que sirve para derivar un conjunto de factores latentes y para caracterizar a los usuarios e ítems con vectores de estos factores, los vectores de ítems más similares a un vector de un usuario representan las recomendaciones. En [42] se usa un modelo de factor latente que mapea usuarios e ítems en un espacio latente de k dimensiones para modelar calificaciones de múltiples dominios. En [50] se genera un modelo híbrido de factorización de matrices con el algoritmo de descenso de gradiente estocástico para usar varias características latentes. En [34] se integra la factorización de matrices probabilística con un modelo

de análisis de texto de una red neuronal profunda. En total 8 estudios usan métodos de factorización de matrices.

El procesamiento del lenguaje natural es el cuarto método más utilizado que comprende y genera el lenguaje natural entre humanos y computadoras [70]. En [42] se integran como documentos las reseñas y las descripciones de aplicaciones multiplataforma, se elimina el ruido al remover palabras vacías y palabras que no están en el idioma inglés, se normalizan los verbos y adjetivos con el método de truncar de Poter y se utiliza el método de LDA para extraer tópicos. En [47] se extraen aspectos identificando referencias hacia aspectos de ítems en reseñas textuales de usuarios, se identifica la polaridad de las opiniones clasificando la orientación de los sentimientos como positivos, neutros o negativos hacia las opiniones de los aspectos identificados en las reseñas y se utilizan métodos basados en vocabulario, en frecuencia de palabras, en relaciones sintácticas y basados en modelos de tópicos. En total 8 estudios utilizan el procesamiento del lenguaje natural.

La optimización es el proceso de buscar el mejor diseño de un sistema sujeto a un conjunto de reglas e identifica el punto que maximiza o minimiza el objetivo de una función [71]. En [46] se genera un modelo de aprendizaje con una función de perdida secuencial llamada en ingles Bayesian Personalized Ranking (BPR) que usa hiper parámetros optimizados con el descenso del gradiente estocástico abreviado en ingles SGD. En [66] se utilizan tres algoritmos de inteligencia de enjambre para agrupar a los usuarios, estos algoritmos son Mussels wandering, colonia de hormigas y búsqueda de Cuckoo. En total 7 estudios usan algoritmos de optimización.

Los algoritmos de vecinos cercanos utilizan como fuente de información una matriz de calificaciones de usuarios por ítems, una función de similitud para identificar usuarios o ítems similares conocidos como vecinos cercanos y con base en los vecinos cercanos se predicen las calificaciones de los ítems que no ha visto del usuario objetivo. En [47] los patrones colaborativos en datos de opiniones de aspectos se explotan con la adaptación de un algoritmo de vecinos cercanos utilizando un enfoque híbrido colaborativo a través del filtrado basado en contenido propuesto por Pazzani. En [38] se genera un filtrado colaborativo de línea de base con el algoritmo de vecinos cercanos basado en el ítem utilizando la similitud del coseno ajustado. En total 6 estudios utilizan algoritmos de vecinos cercanos.

Las redes neuronales artificiales imitan la forma en que las neuronas biológicas se transmiten entre sí en el cerebro humano y son fundamentales en el aprendizaje profundo. En [51] se usa como una función de ponderación una red neuronal prealimentada de dos capas, la capa de entrada toma las características de cada ítem que es un artículo de investigación y la capa de salida contiene un nodo que genera la ponderación. En [67] se usa una red neuronal convolucional para procesar las reseñas

de los ítems y obtener las características semánticas de los ítems. En total 6 estudios usan redes neuronales.

La recuperación de información busca documentos de texto que satisfacen una necesidad de información en grandes colecciones de documentos [72]. En [51] se representa un artículo de investigación en un espacio de palabras con el uso de vectores del tipo tf-idf que contienen valores de palabras y bigramas del título del artículo, el resumen y las palabras clave, además se usa la similitud del coseno para comparar los vectores. En [46] también se usa el método tf-idf y la similitud del coseno en el filtrado basado en contenido porque es un método ampliamente usado para buscar ítems similares a un ítem en particular y porque clasifica los ítems similares. En total 5 estudios usan técnicas de recuperación de información.

La distancia euclidiana es la distancia entre dos puntos en un espacio euclidiano y se calcula con el teorema de Pitágoras. En [51] se utiliza una función de perdida BPR y se preserva la similitud entre ítems del historial de búsqueda con la distancia euclidiana. En [61] se utiliza la distancia euclidiana para calcular la distancia entre vectores de características para el filtrado basado en contenido.

3.7.5 Dominios de aplicación

La distribución de los dominios de aplicación para los enfoques híbridos de recomendación se muestra en la Figura 11, en la cual se demuestra que los dominios de aplicación más usados son las películas, los multidominios, la música, los productos, el turismo, las redes sociales, la ubicación, las aplicaciones y otros dominios como los hospitales, las acciones de sesiones, los trabajos, los artículos de investigación, los nombres personales, los celulares, los médicos y los libros. En total se identificaron 16 dominios de aplicación y se clasificaron en 9 principales tendencias sobre estos dominios de aplicación para enfoques híbridos de recomendación.

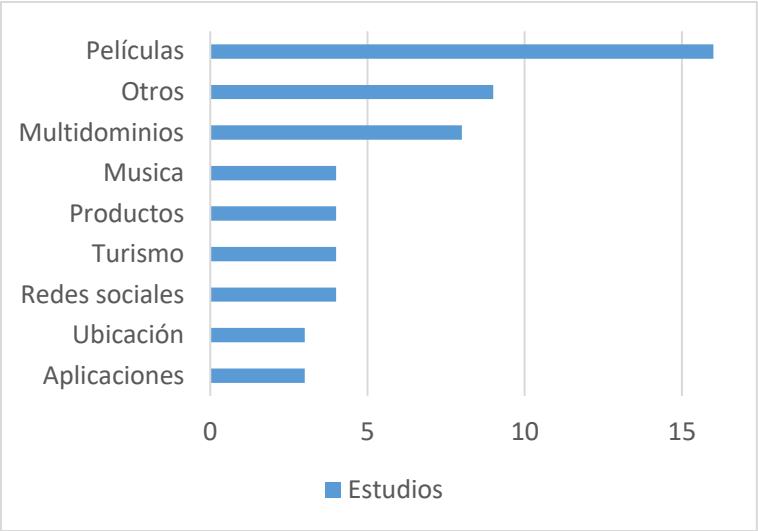

Figura 11. Distribución de los dominios de aplicación para los enfoques híbridos de recomendación.

La tendencia principal es el dominio de películas porque las compañías como Netflix, Amazon, GroupLens, IMDb y otros grupos han generado diversos conjuntos de datos en el dominio de películas y series, en total 16 estudios usan este dominio. La segunda tendencia son los multidominios un gran desafío para los sistemas de recomendación por la integración de diversas fuentes de información y diferentes características de los ítems, sin embargo, Amazon provee conjuntos de datos sobre sus productos clasificados en diversos dominios de aplicación, en total 8 estudios usan los multidominios. Los dominios de aplicación de música, productos, turismo, redes sociales, ubicación y aplicaciones también son utilizados en 3 o más estudios. Otros dominios más específicos son usados en 9 estudios.

3.7.6 Conjuntos de datos

La distribución de los conjuntos de datos se presenta en la Figura 12, esta figura muestra 7 tendencias principales en los conjuntos de datos para enfoques híbridos de recomendación. En total, se identifican 17 conjuntos de datos y se clasifican en 7 tendencias principales. Estas tendencias son MovieLens, conjuntos de datos independientes, Amazon, TripAdvisor, Yelp, Last.FM y otros conjuntos de datos como IMDB, RecSys, Kaggle, Epiones, Douban, ScienceDirect, FilmTrust, Nameling Discovery, Million playlist, Yahoo! y Book crossing.

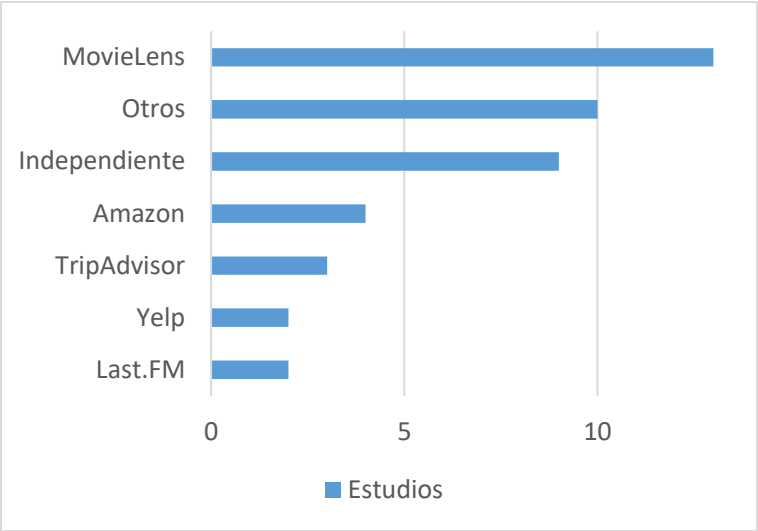

Figura 12. Distribución de los conjuntos de datos para los enfoques híbridos de recomendación.

La primera tendencia es el conjunto de datos de MovieLens, que se utiliza ampliamente en algunas investigaciones debido a que contiene metadatos de películas, etiquetas de películas y valoraciones de usuarios. Además, este conjunto de datos tiene múltiples versiones para varios tipos de investigación. En total, 13 estudios utilizan el conjunto de datos de MovieLens.

La segunda tendencia son los conjuntos de datos independientes, estos conjuntos de datos se desarrollan en sus propias investigaciones. En total, 9 estudios utilizan conjuntos de datos independientes.

La tercera tendencia es el conjunto de datos de Amazon, que es un conjunto de datos multidominio porque tiene varios tipos de productos comerciales como ropa, películas, joyas, libros, etc. Además, contiene metadatos de los artículos, valoraciones de usuarios y reseñas de texto de los usuarios. En total, 4 estudios utilizan el conjunto de datos de Amazon.

La cuarta tendencia es el conjunto de datos de TripAdvisor, que se utiliza ampliamente en el ámbito del turismo. En total, 3 estudios utilizan el conjunto de datos de TripAdvisor.

La quinta y sexta tendencia son el conjunto de datos de Yelp (multidominio) y el conjunto de datos de Last.FM (dominio musical). En total, cada conjunto de datos es utilizado por 2 estudios. Otros conjuntos de datos, como IMDB, RecSys, Kaggle, etc., son utilizados por 10 estudios.

Las tendencias de conjuntos de datos previamente mencionadas se deben a las siguientes características que son la frecuencia de uso del conjunto de datos en otras

investigaciones, la fuente de información que se utilizó para desarrollar el conjunto de datos, el tamaño del conjunto de datos, el dominio de aplicación y el tipo de datos que presenta como calificaciones, metadatos de los ítems, datos demográficos, reseñas textuales, datos de comportamiento e interacción, datos de redes sociales, etc.

3.7.6 Métricas de evaluación

La distribución de las métricas de evaluación para los enfoques híbridos de recomendación se muestra en la Figura 13, en la cual se demuestra que las métricas de evaluación más utilizadas son recall con 16 estudios, precision con 14, RMSE con 11, NDCG con 8, F1 con 7, MAE con 7, accuracy con 5, las pruebas de usuario con 5, coverage con 3, MAP con 3 y otras métricas como novelty, crowd-sourced, ROC, silhoutte, rate coverage, la similitud, mean reciprocal rank, MSE, user space coverage, aggregate diversity, expected popularity complement, rango de clics, RMSS, R-precision, cold start y NMAE, estas otras métricas con 17 estudios. En total se identificaron 26 métricas de evaluación y se clasificaron en 11 principales tendencias sobre métricas de evaluación para enfoques híbridos de recomendación.

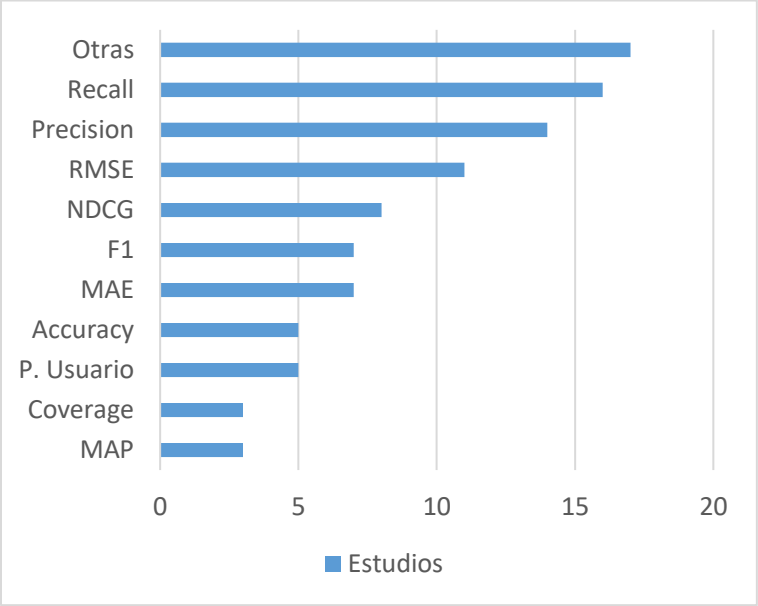

Figura 13. Distribución de las métricas de evaluación para los enfoques híbridos de recomendación.

La principal tendencia son las evaluaciones fuera de línea en inglés offline que se realizan sobre conjuntos de datos que son particionados generalmente 80% para entrenamiento y 20% para pruebas, además se utiliza generalmente la validación cruzada con 5 pliegos. Las evaluaciones en línea en inglés online se realizan a usuarios reales con pruebas de usuario y se usa el crowdsourcing como Amazon Mechanical

Turk (MTurk) [36], las evaluaciones online también son una oportunidad para nuevas investigaciones a pesar de que pocos estudios evalúan de esta forma sus experimentos por el costo y el tiempo de la evaluación.

3.8 Conclusiones

En la actualidad, los sistemas de recomendación representan un alto impacto económico, social y tecnológico a nivel internacional, debido a que las principales empresas tecnológicas han utilizado estos sistemas en sus servicios líderes. Los sistemas de recomendación han contribuido a resolver el problema de la sobrecarga de información, mejorar la experiencia del usuario, la toma de decisiones del usuario y las ventas de las empresas. Además, los enfoques híbridos se utilizan ampliamente para optimizar los sistemas de recomendación.

Se ha realizado una revisión sistemática de la literatura sobre los enfoques híbridos para sistemas de recomendación con el fin de desarrollar un estado del arte desde 2016 hasta 2020, considerando una brecha de conocimiento de un trabajo relacionado principal. Además, a diferencia de otros trabajos, esta investigación utiliza una guía más reciente dirigida a los sistemas de información, que es un área principal relacionada con los sistemas de recomendación.

A continuación, se presentan los hallazgos principales identificados en los enfoques híbridos para sistemas de recomendación. Las tendencias en problemas incluyen el arranque en frío, la exactitud, la escasez de datos, la integración de fuentes de datos, la escalabilidad y otros problemas. Las tendencias en técnicas de recomendación son el filtrado colaborativo, el filtrado basado en contenido, el filtrado basado en conocimiento, el filtrado basado en secuencias, el filtrado demográfico, el filtrado contextual y otras técnicas. Las tendencias en enfoques híbridos son ponderados, mixtos, de selección, combinación de características, aumento de características, en cascada y otros enfoques híbridos. Las tendencias en métodos y algoritmos incluyen el agrupamiento, la similitud del coseno, la factorización de matrices, el procesamiento del lenguaje natural, algoritmos de optimización, algoritmos de los k-vecinos más cercanos, redes neuronales, recuperación de información, distancia euclidiana y otros métodos y algoritmos. Las tendencias en dominios de aplicación son películas, multi-dominios, música, productos, turismo, redes sociales, ubicación, aplicaciones y otros dominios. Las tendencias en conjuntos de datos son MovieLens, conjuntos de datos independientes, Amazon, TripAdvisor, Yelp, Last.FM y otros conjuntos de datos. Las tendencias en métricas de evaluación incluyen recall, precisión, RMSE, NDCG, F1, MAE, exactitud, pruebas de usuario, cobertura, MAP y otras métricas de evaluación.

Estos hallazgos sobre desafíos, metodologías, conjuntos de datos, dominios de aplicación y métricas de evaluación en enfoques híbridos beneficiarán a la comunidad de sistemas de recomendación para generar nuevas investigaciones e identificar nuevas áreas de oportunidad en los enfoques híbridos de recomendación.

Los trabajos futuros incluyen ampliar el estado del arte con años posteriores a 2020, refinar los términos de búsqueda, incluir otras bases de datos científicas, mejorar los pasos prácticos de selección y evaluación de calidad, buscar una guía más reciente que aborde los sistemas de recomendación, y seleccionar un dominio de aplicación específico.

Agradecimientos

Los autores agradecen al Laboratorio de Ingeniería del Lenguaje y del Conocimiento (LKE) de la Benemérita Universidad Autónoma de Puebla y al Consejo Nacional de Ciencia y Tecnología (CONACYT).

Capítulo 4: Un modelo recomendación híbrido basado en RI para textos turísticos mexicanos en Rest-Mex 2022

La Organización Mundial del Turismo (OMT) define el concepto de turismo como "un fenómeno social, cultural y económico que implica el movimiento de personas a países o lugares fuera de su entorno habitual por motivos personales o profesionales. Estas personas se llaman visitantes (turistas, excursionistas, residentes o no residentes) y el turismo tiene que ver con sus actividades, algunas de las cuales implican gastos turísticos" [73]. Hoy en día, los volúmenes de las exportaciones de petróleo, productos alimenticios y automóviles han sido superados por el turismo. Por esta razón, el turismo es uno de los principales sectores económicos en el mundo, ya que esta actividad mejora las exportaciones, aumenta el número de empleos y desarrolla la economía [74]. En México, el turismo representa el 8.7% del producto interno bruto (PIB) nacional y genera 4.5 millones de empleos directos. Sin embargo, este sector económico ha sido afectado por la pandemia de COVID-19, que se extendió en México en marzo de 2020. Por lo tanto, se debe mejorar la calidad y seguridad de los productos y servicios turísticos para restaurar el turismo mexicano [75].

Así, surge el foro internacional llamado Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts (Rest-Mex) [76] del Iberian Languages Evaluation Forum (IberLEF), el cual tiene el objetivo de fortalecer el turismo mexicano a través de novedosos mecanismos de procesamiento de lenguaje natural. Mientras que el IberLEF es una campaña de evaluación compartida de sistemas de PLN en español y otras lenguas ibéricas que se organiza desde 2019, y se celebra en el marco del congreso anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN).

El Rest-Mex 2022 presentó las siguientes tres tareas compartidas: (1) Sistema de recomendación, (2) Análisis de sentimientos y (3) Semáforo epidemiológico. La tarea de sistema de recomendación es abordada por esta tesis, y se define de la siguiente manera: "*Dado un turista de TripAdvisor y un lugar turístico mexicano, el objetivo es obtener*

automáticamente el grado de satisfacción (entre 1 y 5) que el turista tendrá al visitar ese lugar." La motivación de esta tarea es que pocos sistemas de recomendación para sitios turísticos se basan en la afinidad del perfil de un usuario en comparación con la descripción de cada lugar. Las colecciones de datos para entrenar estos sistemas provienen de usuarios y lugares en países de habla inglesa. Considerando la importancia de los países iberoamericanos en el turismo, es de vital importancia generar recursos en español que permitan la creación de sistemas que ayuden a desarrollar sistemas inteligentes en el turismo.

Los sistemas de recomendación (SR) son herramientas de software que sugieren varios productos llamados ítems, como productos comerciales, sitios web, amigos, empleos, películas, canciones, lugares turísticos, hoteles, restaurantes y otros ítems, basándose en las preferencias de los usuarios [31]. Hoy en día, los SR representan un alto impacto económico, social y tecnológico a nivel internacional, ya que las principales compañías como Google, Facebook, Twitter, LinkedIn, Netflix, Amazon, Microsoft, Yahoo!, eBay, Pandora, Spotify y otras más utilizan estos sistemas en sus servicios líderes [2]. Además, estas compañías se benefician económicamente de los SR porque abordan el problema de la sobrecarga de información en el comercio electrónico, mejoran la experiencia del usuario y facilitan la toma de decisiones de los usuarios [32]. Por otro lado, el rendimiento de los SR puede optimizarse con enfoques híbridos que combinan dos o más algoritmos de recomendación para complementar sus desventajas con las ventajas de otros algoritmos. Por esta razón, se invierten grandes sumas de dinero para optimizar algoritmos y desarrollar nuevas investigaciones en enfoques híbridos [77].

La recuperación de información (RI) busca documentos de texto en grandes colecciones de documentos basándose en una necesidad de información [72]. Las técnicas de RI se han utilizado ampliamente por la técnica de recomendación llamada filtrado basado en contenido (CBF, por sus siglas en inglés), que recomienda ítems similares [65]. En esta técnica, se genera un perfil de usuario a partir de los atributos y descripciones textuales de los ítems valorados por el usuario, luego se recomiendan ítems similares al perfil del usuario; además, esta técnica utiliza el metadato textual de los ítems como fuente principal de información [78]. Aunque las técnicas de RI se han utilizado tradicionalmente en el CBF, también pueden ser empleadas por otra técnica de recomendación llamada filtrado colaborativo (CF, por sus siglas en inglés), que utiliza una matriz de valoraciones usuario-ítem como fuente principal de información para buscar similitudes entre usuarios (CF usuario-usuario) o similitudes entre ítems (CF ítem-ítem) llamados vecinos cercanos. Además, las valoraciones de los usuarios suelen ser valores numéricos entre 1 y 5, y las predicciones de valoraciones de los usuarios se realizan con los vecinos más cercanos identificados [40].

En este capítulo se presenta un modelo híbrido de recomendación basado en recuperación de información para la tarea de sistema de recomendación de Rest-Mex 2022, donde se aplicaron las técnicas de recomendación CF ítem-ítem y CBF utilizando técnicas de recuperación de información. La estructura de este capítulo es la siguiente. Los trabajos relacionados se presentan en la sección 4.1, el modelo propuesto se expone

en la sección 4.2, los experimentos y resultados se muestran en la sección 4.3 y las conclusiones se describen en la sección 4.4.

4.1 Trabajos previos de la tarea de sistema de recomendación del Rest-Mex 2021

La tarea de recomendación en la edición Rest-Mex 2021 [75] consistió en predecir el grado de satisfacción que un turista tendrá al visitar un lugar en Nayarit, México. El conjunto de datos cuenta con 2,263 instancias, incluyendo 2,011 turistas y 18 lugares turísticos de Nayarit, y su información proviene de TripAdvisor. El conjunto de datos se dividió en 70/30, lo que significa que el conjunto de entrenamiento contiene 1,587 instancias y el conjunto de prueba contiene 681 instancias. Además, se utilizó la métrica de error medio absoluto (MAE) para la evaluación de la competencia.

En esa edición del Rest-Mex, participaron dos equipos: Alumni-MCE 2GEN y Labsemco-UAEM. La Tabla 10 describe los resultados del Rest-Mex 2021, donde los mejores resultados, con 0.31 y 0.32 de MAE, son presentados por el equipo Alumni-MCE 2GEN, mientras que el equipo Labsemco-UAEM presenta un MAE de 1.65.

Tabla 10. Resultados obtenidos en el Rest-Mex 2021 [75].

Team	MAE
Alumni-MCE 2GENRun1	0.31
Alumni-MCE 2GENRun2	0.32
Baseline	0.73
Labsemco-UAEM	1.65

El equipo Alumni-MCE 2GEN presentó una investigación titulada *An Embeddings Based Recommendation System for Mexican Tourism. Submission to the REST-MEX Shared Task at IberLEF 2021* [79], que propone dos variantes del mismo modelo, donde el cambio más relevante es una representación vectorial de palabras. El conjunto de datos se preprocesa para limpiar y obtener información completamente en español, luego se aplica un modelo Doc2Vec a la información de los usuarios y de los lugares. Además, se diseña una matriz con los embeddings obtenidos y se utiliza una red neuronal con una capa oculta en la primera variante del modelo. En la segunda variante del modelo, se propone un sistema basado en representaciones distribuidas para textos.

El equipo Labsemco-UAEM presentó una investigación titulada *A Recommendation System for Tourism Based on Semantic Representations and Statistical Relational Learning* [80], que propone un método de representación de texto diferente de los métodos de co-ocurrencia léxica en el texto. Este método extrae las características lingüísticas del

texto, específicamente las señales léxicas y semánticas de sinonimia y antonimia. Además, se utiliza el modelo ComplEx para generar recomendaciones basadas en las relaciones entre los usuarios y los lugares.

Estos trabajos relacionados abordan principalmente el problema de precisión de los sistemas de recomendación, ya que es el principal desafío en la competencia del Rest-Mex. Sin embargo, el problema de la escalabilidad es otro desafío principal para los SR en la era digital, debido a que los motores de recomendación deben procesar grandes volúmenes de información con tiempos de ejecución mínimos [68]. Por esta razón, en esta investigación se utilizan técnicas de RI y estructuras de datos de índice invertido para reducir el costo computacional de las recomendaciones.

4.2 Metodología

El conjunto de datos, las técnicas de recuperación de información y las técnicas de sistemas de recomendación utilizadas para desarrollar un modelo híbrido de recomendación basado en la recuperación de información para textos turísticos mexicanos en Rest-Mex 2022 se presentan en esta sección. El diseño de la metodología se muestra en la Figura 14, que contiene cinco componentes principales: el conjunto de datos, el preprocesamiento, las técnicas de recuperación de información, las técnicas de sistemas de recomendación y la evaluación. A continuación, se describen cada uno de estos componentes.

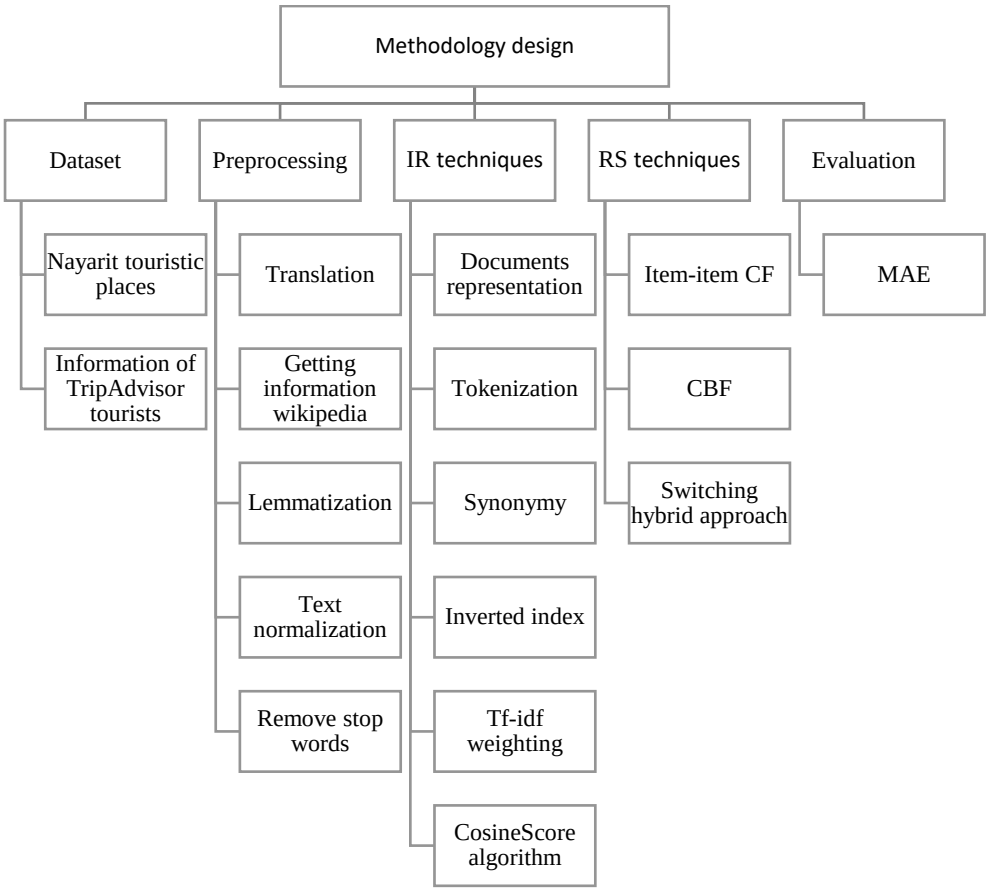

Figura 14. Diseño de la metodología.

4.2.1 Conjunto de datos

El conjunto de datos tiene 2,263 instancias con 2,011 turistas y 18 lugares turísticos de Nayarit, México. Este conjunto de datos se obtuvo de los turistas que compartieron su grado de satisfacción en TripAdvisor entre 2010 y 2020. Cada clase de satisfacción es un entero entre 1 y 5, donde 1 es muy malo, 2 es malo, 3 es neutral, 4 es bueno y 5 es muy bueno. El objetivo de la tarea del sistema de recomendación de Rest-Mex 2022 es predecir la clase de satisfacción que un turista puede tener al recomendar un destino.

La distribución de las clases en el conjunto de datos de entrenamiento es la siguiente: la clase 1 tiene 45 instancias, la clase 2 tiene 53 instancias, la clase 3 tiene 167 instancias, la clase 4 tiene 457 instancias y la clase 5 tiene 860 instancias. El conjunto de datos de entrenamiento tiene un total de 1,582 instancias, que representan alrededor del 70 % del total de instancias. La distribución de las clases en el conjunto de datos de prueba es la siguiente: la clase 1 tiene 20 instancias, la clase 2 tiene 24 instancias, la clase 3 tiene 72 instancias, la clase 4 tiene 196 instancias y la clase 5 tiene 369 instancias. El conjunto de datos de prueba tiene un total de 681 instancias, que representan alrededor

del 30 % del total de instancias. El desequilibrio de clases representó un gran desafío para esta tarea del Rest-Mex 2022.

Cada instancia consta de dos partes de información. (1) La información del usuario que contiene el género del turista, el lugar turístico que el turista recomienda visitar, el lugar de origen del turista, la fecha de la recomendación, el tipo de viaje (familia, amigos, solo, pareja y negocios) y el historial de los lugares que el turista ha visitado con sus opiniones o reseñas sobre cada uno de estos lugares. (2) La información del lugar contiene una breve descripción textual del lugar y un conjunto de características como el tipo de turismo (aventura, playa, relajación, entre otros), el tipo de ambiente (familiar, privado o público), si es gratuito o de pago, entre otros [75].

4.2.2 Preprocesamiento

La biblioteca langdetect de Python se utiliza para identificar las reseñas de los usuarios en inglés, luego la biblioteca googletrans se usa para traducir las reseñas de los usuarios en inglés al español en el primer paso de preprocesamiento. La biblioteca Wikipedia se emplea para obtener información sobre los lugares relacionados con las reseñas de los usuarios en el segundo paso de preprocesamiento, ya que el conjunto de datos solo contiene información sobre los 18 lugares de Nayarit, México. La biblioteca Spacy se usa para lematizar el conjunto de datos en el tercer paso de preprocesamiento.

La normalización del texto se aplica para convertir las letras mayúsculas en minúsculas, eliminar acentos y eliminar signos de puntuación como [,.;:#$!¡?%\n\f=*@|] en el cuarto paso de preprocesamiento. Se utiliza una lista de palabras vacías de la biblioteca Spacy para eliminar palabras vacías en el conjunto de datos en el quinto paso de preprocesamiento.

El resultado del preprocesamiento permite tokenizar los documentos del conjunto de datos de una manera más eficiente para seleccionar los términos de cada documento de una forma estandarizada y desarrollar un índice invertido como estructura de datos para las técnicas de recuperación de información.

4.2.3 Técnicas de recuperación de información

Se genera un modelo de espacio vectorial, ya que este modelo representa un conjunto de documentos como vectores en un espacio vectorial común y cada término de los documentos representa una dimensión del espacio vectorial [72]. Además, este modelo es fundamental para varias operaciones de recuperación de información, como la puntuación de documentos en una consulta, la clasificación de documentos y el agrupamiento de documentos. Asimismo, la representación de la consulta como vector

es muy importante para encontrar los vectores de documentos más similares al vector de la consulta utilizando la similitud del coseno. La fórmula de la similitud del coseno es la siguiente.

$$sim(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{(\vec{V}(d_1) \, ||\vec{V}(d_2))},$$

Donde la similitud entre dos documentos d_1 y d_2 se calcula con la similitud del coseno y su representación vectorial $\vec{V}(d_1)$ y $\vec{V}(d_2)$, además el numerador representa el producto punto de los vectores, mientras que el denominador es el producto de sus longitudes euclidianas. El esquema de ponderación de frecuencia de término - frecuencia inversa de documento (tf-idf) se utiliza para asignar a un término t un peso en el documento d dado por la siguiente fórmula.

$$tf - idf_{t,d} = tf_{t,d} \times idf_t$$

Donde el peso es (1) mayor cuando t ocurre muchas veces dentro de un pequeño número de documentos, (2) es menor cuando el término ocurre pocas veces en un documento u ocurre en muchos documentos y (3) es el menor cuando el término ocurre en prácticamente todos los documentos. La frecuencia del término tf y la frecuencia inversa de documentos $idf_{t,d}$ se normalizan con la función logarítmica mediante las siguientes fórmulas.

$$tf_{t,d} = 1 + \log(tf_{t,d})$$

$$idf_t = \log\left(\frac{N}{df_t}\right)$$

Un modelo de espacio vectorial con esquema de ponderación tf-idf y similitud del coseno se implementa en Python utilizando el algoritmo CosineScore de Manning [72]. En este modelo, los lugares turísticos se representan como consultas y las reseñas de los usuarios se representan como documentos. Luego, los términos de los documentos de las reseñas de los usuarios se obtienen mediante el proceso de tokenización y estos términos se indexan en una estructura de datos de índice invertido. Además, se utiliza un diccionario de sinónimos para indexar los sinónimos de los términos en el índice invertido.

4.2.4 Modelo de recomendación híbrido basado en RI

Se genera un modelo híbrido de recomendación aplicando las técnicas de recomendación de filtrado colaborativo ítem-ítem, filtrado basado en contenido y enfoque híbrido por selección. Este modelo se describe en la Figura 15, donde el

enfoque híbrido por selección se utiliza de la siguiente manera: el filtrado colaborativo ítem-ítem se emplea para usuarios que contienen suficiente información en sus reseñas sobre lugares turísticos, mientras que se utiliza el filtrado basado en contenido para usuarios que no tienen reseñas, abordando así el desafío del arranque en frío.

Se desarrolla un índice invertido para cada usuario en el filtrado colaborativo ítem-ítem, mientras que se desarrolla un índice invertido con todos los documentos de las reseñas de los usuarios en el filtrado basado en contenido. Ambas técnicas utilizan un documento del lugar turístico recomendado a un turista como consulta, luego se obtienen los k documentos de reseñas de turistas más similares a la consulta para promediar sus calificaciones y generar el grado de satisfacción (entre 1 y 5) que el turista tendrá al visitar ese lugar. Finalmente, las predicciones de las recomendaciones se evalúan mediante la métrica MAE.

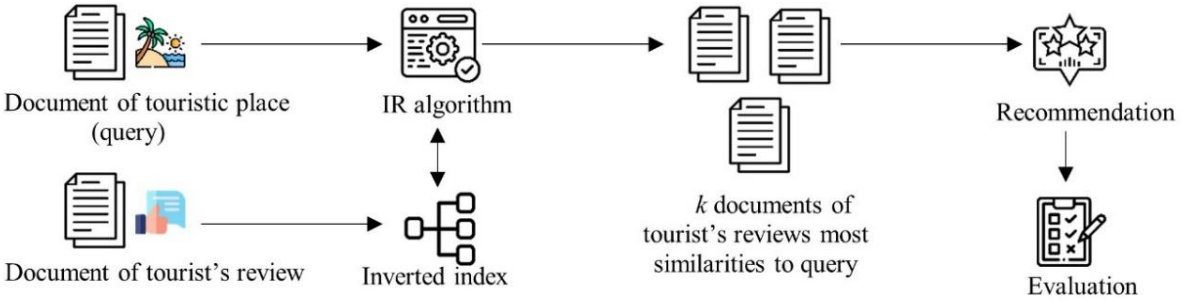

Figura 15. Modelo recomendación híbrido basado en RI para el turismo [81].

4.3 Experimentos y resultados

El primer experimento se realiza con el conjunto de datos de entrenamiento, que tiene 1,582 instancias, representando alrededor del 70 % del total de instancias. El modelo híbrido de recomendación basado en recuperación de información se utiliza en cada instancia para generar su predicción. Los resultados del modelo híbrido de recomendación se evalúan mediante la métrica MAE, que fue seleccionada por Rest-Mex 2022. Además, esta métrica se emplea comúnmente cuando hay muchos valores atípicos y mide la distancia entre el vector de predicciones y el vector de valores objetivo [82]. La fórmula del MAE es la siguiente [83].

$$MAE = \frac{\sum_{u \in U} \sum_{i \in testset_u} |rec(u,i) - r_{u,i}|}{\sum_{u \in U} |testset_u|}$$

Donde MAE calcula la desviación promedio entre las puntuaciones de recomendación computadas $rec(u,i)$ y el valor real de la calificación $r_{u,i}$ para todos los usuarios evaluados $u \in U$ y todos los ítems en sus conjuntos de prueba $testset_u$.

Se realizaron cuatrocientos pruebas para identificar los mejores valores para los parámetros *k1* y *k2,* que corresponden a los *k* documentos de reseñas de turistas más similares a la consulta para el algoritmo de IR. El parámetro k1 se utiliza para el filtrado basado en contenido (CBF) y el parámetro *k2*. se utiliza para el filtrado colaborativo ítem-ítem (CF). Los tres mejores resultados de este experimento se presentan en la Tabla 11.

Tabla 11. Los tres mejores resultados para el primer experimento en el conjunto de entrenamiento.

Test number	K1	K2	MAE
1	14	13	0.7111251580278128
2	15	18	0.7117572692793932
3	13	18	0.7117572692793932

El segundo experimento se realiza con el conjunto de datos de prueba, que tiene 681 instancias, representando alrededor del 30 % del total de instancias. El modelo híbrido de recomendación basado en recuperación de información se utiliza con los parámetros identificados en los tres mejores resultados del primer experimento con el conjunto de datos de entrenamiento. Luego, se envían tres salidas a Rest-Mex 2022, cuyos resultados se presentan en la Tabla 3. Donde, MAE es la métrica principal para la evaluación de la competencia y se utilizaron otras métricas como accuracy, f-measure, macro recall y macro precision.

El primer resultado con *k1=13* y *k2=18* obtuvo un MAE de 0.6981707317 con el cual se logró el segundo lugar en la tarea de sistema de recomendación del Rest-Mex 2022, con una diferencia de 0.004814684 de MAE respecto al primer lugar. El segundo resultado con *k1=14* y *k2=18* obtuvo un MAE de 0.699695122, que representa el tercer lugar en la competencia y se obtuvo un macro recall de 0.2548259978 que es el mejor resultado de esa métrica de evaluación en el Rest-Mex [76].

Tabla 12. Los resultados para el segundo experimento sobre el conjunto de datos de prueba.

Team	MAE	Accuracy	F-measure	Macro Recall	Macro Precision
LKEBUAP_RUN_k1_13_k2_18	0.6981707317	46.64634146	0.2215145883	0.242524662	0.2303778647
LKEBUAP_RUN_k1_14_k2_18	0.699695122	45.88414634	0.2196485875	0.2548259978	0.2280366323

4.4 Conclusiones

Esta investigación presentó un modelo híbrido de recomendación basado en la recuperación de información para textos turísticos mexicanos en el Rest-Mex 2022. En este modelo, se implementó un modelo de espacio vectorial con esquema de ponderación tf-idf y similitud del coseno utilizando Python. Además, se generó un modelo híbrido de recomendación aplicando las técnicas de recomendación de filtrado colaborativo ítem-ítem, filtrado basado en contenido y un enfoque híbrido por selección. Luego, se realizaron dos experimentos. El primero utilizó el conjunto de datos de entrenamiento para identificar los mejores parámetros para el algoritmo de recuperación de información. El segundo experimento utilizó el conjunto de datos de prueba con los tres mejores parámetros identificados en el primer experimento para enviar tres listas de resultados a la tarea compartida, con las cuales se logró el segundo y tercer lugar en la tarea de sistema de recomendación del Rest-Mex 2022. Y se alcanzó el mejor resultado de la métrica de macro recall. Finalmente, este trabajo es una aportación para la comunidad de sistemas de recomendación que tiene el objetivo de beneficiar al turismo mexicano.

Capítulo 5: Equipo LKE-IIMAS en Rest-Mex 2023: Análisis de sentimientos en reseñas turísticas mexicanas usando una adaptación de dominio basada en Transformer

Esta sección describe la participación del equipo LKE-IIMAS en la tarea de análisis de sentimiento del Rest-Mex 2023 [84]. El análisis de sentimiento en textos turísticos ha ganado relevancia en la última década porque el turismo es un fenómeno social, cultural y económico que contribuye en el desarrollo económico de múltiples países. Por ejemplo, en México esta actividad ha representado el 8.7% del producto interno bruto (PIB), ha generado 4.5 millones de trabajos directos [76]. En consecuencia, nuevos mecanismos de procesamiento de lenguaje natural son requeridos para mejorar los servicios turísticos y conocer las necesidades de los turistas. Esta tarea se centra en el idioma español debido a que la mayoría de los trabajos relevantes del estado del arte se realizan en idioma inglés, y solo algunos estudios han utilizado únicamente el idioma español peninsular. Esta tarea compartida proporciona una colección de textos recopilados de opiniones de turistas [85].

La tarea compartida de análisis de sentimiento del Rest-Mex 2023 consiste en un problema de clasificación de texto donde los sistemas participantes deben predecir la polaridad de una opinión generada por un turista quien viajo a los lugares, restaurantes y hoteles más representativos en México, Cuba, y Colombia. Esta colección de datos fue obtenida de los turistas quienes compartieron su opinión en TripAdvisor entre el 2002 y el 2022. Cada clase de la opinión es representada como un numero entero entre 1 y 5, donde 1 representa la polaridad más negativa y 5 es la polaridad más positiva. Así mismo, cada opinión tiene un tipo de etiqueta por cada clase. El problema se define de la siguiente forma: "Dado una opinión acerca de un lugar turístico de México, el objetivo es determinar la polaridad entre 1 y 5 del texto, el tipo de opinión (hotel,

restaurante o atracción) y el país del lugar del cual la opinión ha sido dada (México, Cuba o Colombia)" [84].

La complejidad de los idiomas dificultó la clasificación de textos irónicos, sarcásticos y subjetivos en análisis de los sentimientos [86]. Adicionalmente, los modelos basados en redes neuronales recurrentes del tipo LSTM (Long-short term memory), los enfoques de fastText, las redes neuronales convolucionales (CNN), y las arquitecturas transformers han sido usados como técnicas de aprendizaje automático (ML) en la literatura [86] en donde los enfoques de fastText han sido más rápidos que otros modelos de ML. Las arquitecturas transformers han logrado un alto rendimiento en sus predicciones a diferencia de otros modelos de ML. Además, los mejores resultados en ediciones previas del Rest-Mex fueron logrados con enfoques basado en transformers [76] [75].

Por esta razón, se propone una adaptación de dominio basada en transformer para mejorar la precisión de la clasificación de los sentimientos en las reseñas turísticas. Esta sección esta organizada de la siguiente forma. En la subsección 5.1 se presentan los trabajos relacionados. En la subsección 5.2, se describen los fundamentos sobre transformers. En la subsección 5.3, se analizan los datos. En la subsección 5.4, se explica la metodología propuesta en detalle. En la subsección 5.5, se muestran los resultados y en la subsección 5.6, se presentan las conclusiones.

5.1 Trabajos relacionados

Las ediciones anteriores de Rest-Mex abarcaron una diversa gama de métodos para la clasificación del análisis de sentimientos. Estos métodos incluyeron modelos basados en transformers, como versiones ajustadas de modelos tipo Bert, clasificadores de regresión logística, el algoritmo Naive Bayes Multinomial, una cascada de clasificadores binarios, técnicas bayesianas, KNN con distancia de Jaccard y un enfoque basado en LDA para la extracción de temas. Además, se empleó una arquitectura simple de aprendizaje profundo para la clasificación.

En Rest-Mex 2021 [75], el equipo que logró el mejor enfoque utilizó dos estrategias basadas en Bert [87]. La estrategia inicial consistió en ajustar BETO, un modelo preentrenado tipo Bert en español. La segunda estrategia combinó embeddings de Bert con vectores de características ponderados por TF-IDF. El segundo mejor enfoque en 2021 involucró la utilización del modelo BETO y una cascada de clasificadores binarios basados en él [88]. El tercer mejor enfoque propuso una técnica de extracción de palabras clave no supervisada para construir una lista de pesos de sentimiento [89]. Utilizaron SVM para la clasificación, empleando representaciones vectoriales de

entidades textuales y emparejando las puntuaciones de palabras prototípicas con las etiquetas de los textos.

Para la campaña de evaluación de Rest-Mex 2022, los organizadores utilizaron una nueva métrica de evaluación para valorar la tarea de análisis de sentimientos de esa edición en particular. Esta métrica se define en la ecuación 1.

$$measure_S = \frac{\frac{1}{1 + MAE_p} + F_A}{2}$$

Donde F_A es el promedio entre la micro F-medida para cada clase, y MAE evalúa la polaridad.

$$MAE_{S_x} = \frac{1}{n}\sum_{i=1}^{n}|T(i) - S_x(i)|$$

Donde S_x es un sistema participante x, T(i) es el resultado de la instancia i y $S_x(i)$ es el resultado del sistema participante x, por ejemplo, i. Finalmente, n es el numero de instancias en la colección.

En Rest-Mex 2022 [90], el equipo UMU logró los mejores resultados al abordar el análisis de sentimientos como dos desafíos distintos: la clasificación de polaridad como un problema de regresión y la resolución de objetivos de opinión como un problema de clasificación múltiple. Su pipeline incluyó tareas como limpieza de texto, extracción de características lingüísticas y uso de embeddings de oraciones no contextuales y contextuales utilizando los modelos FastText, BERT y RoBERTa. También realizaron ajuste fino y optimización de hiperparámetros, entrenando modelos de redes neuronales con pérdida RMSE para la regresión y pérdida de entropía cruzada binaria para la clasificación [91]. El equipo que obtuvo el segundo lugar, UC3M, exploró dos enfoques. El primer enfoque implicó el uso de SVM, una técnica tradicional de aprendizaje automático, para la clasificación de texto. El segundo enfoque utilizó transformers ajustados con diferentes modelos preentrenados [88]. El equipo que obtuvo el tercer lugar, CIMAT, siguió una metodología en dos pasos. Obtuvieron características de alto nivel de los textos utilizando un ensamble de modelos BERT entrenados para ambas subtareas. En el segundo paso, crearon un ensamble óptimo de clasificadores entrenados con las características obtenidas en el primer paso para hacer predicciones finales para cada subtarea. Su preprocesamiento incluyó pasos mínimos, como concatenar el título y la opinión y convertir el texto a minúsculas [89].

5.2 Fundamentos

Los transformers se volvieron muy populares en la comunidad de procesamiento de lenguaje natural, propuestos por Vaswani et al. en 2017 [92], esta arquitectura neuronal logró y superó resultados de vanguardia para muchas tareas de NLP. Su mecanismo central es un proceso de autoatención que enfatiza las relaciones contextuales de palabras o tokens.

Estas arquitecturas de transformers abordaron el problema de las dependencias a largo plazo utilizando un mecanismo de autoatención, que permite al modelo ponderar la importancia de diferentes palabras o tokens en la secuencia mientras procesa la secuencia completa simultáneamente. Mediante el mecanismo de autoatención, también conocido en inglés como scaled dot-product attention, los autores permitieron que el modelo calcule la importancia o los pesos de atención para cada palabra/token en la secuencia de entrada basándose en sus relaciones con otras palabras/tokens. La atención permite que el modelo se concentre en las partes relevantes de la entrada durante las etapas de codificación y decodificación.

Desde esta perspectiva, el uso de mecanismos de autoatención en los transformers demuestra tener la capacidad de desempeñarse bien en situaciones de dependencias en el espacio de entrada, tanto localmente como en dependencias a largo plazo. Esta capacidad ha hecho que los transformadores sean altamente efectivos en varias tareas de NLP, incluido el análisis de sentimientos, donde capturar información contextual y dependencias es crucial para una clasificación de sentimientos precisa. Sin embargo, algunas variantes de los transformers merecen ser enfatizadas en el contexto de nuestra tarea: los transformers bidireccionales y los transformers autorregresivos, que son dos enfoques diferentes dentro de la arquitectura de transformers que sirven para propósitos distintos.

Los transformers Bidireccionales, como BERT (Bidirectional Encoder Representations from Transformers), están diseñados para capturar información contextual considerando tanto el contexto izquierdo como derecho de cada palabra en la secuencia de entrada. Logran esto mediante modelado de lenguaje enmascarado, donde algunas palabras en la entrada se enmascaran aleatoriamente, y el modelo se entrena para predecir esas palabras enmascaradas basándose en el contexto circundante. Los modelos basados en BERT se han utilizado ampliamente para diversas tareas de NLP, incluido el análisis de sentimientos, ya que pueden aprovechar efectivamente el contexto bidireccional para entender las relaciones entre las palabras [93].

Por otro lado, los transformers autorregresivos, como GPT (Generative Pre-trained Transformer), se centran en generar una salida coherente y contextualmente relevante procesando secuencialmente la secuencia de entrada, típicamente de izquierda a

derecha. El modelo predice la siguiente palabra en la secuencia durante el entrenamiento basándose en el contexto precedente. Este enfoque de generación secuencial hace que los modelos autorregresivos sean adecuados para tareas de generación y completado de texto. En el análisis de sentimientos, los modelos autorregresivos pueden generar texto relacionado con el sentimiento, resumir sentimientos o generar respuestas basadas en sentimientos dados [94].

Los hechos contextuales son una pieza importante de información para la tarea de análisis de sentimientos. Según la literatura [92] [95], el enfoque tradicional basado en bolsa de palabras o n-gramas ignora la estructura de correlación entre palabras al tratar cada palabra de manera aislada. En otras palabras, ignoran la dependencia entre ellas. Sin embargo, el mecanismo de autoatención de la arquitectura Transformer considera la importancia del contexto alrededor de cada palabra. Este mecanismo de autoatención permite a esta arquitectura capturar dependencias a largo plazo y entender el sentimiento expresado de manera más matizada.

Hay una amplia variedad de situaciones donde los transformadores pueden mejorar el análisis de sentimientos. Aprenden incrustaciones para cada palabra de manera no supervisada. Estas incrustaciones de palabras codifican información semántica y sintáctica, lo que permite al modelo capturar patrones de sentimiento más matizados y generalizar mejor a datos no vistos.

Para el aprendizaje por transferencia, los transformers pueden ser preentrenados en grandes corpus de texto utilizando tareas como modelado de lenguaje enmascarado o predicción de la siguiente oración. Este preentrenamiento permite a los transformers adquirir un entendimiento general del lenguaje, que luego puede ser ajustado finamente en tareas específicas de análisis de sentimientos con conjuntos de datos etiquetados más pequeños. El aprendizaje por transferencia ayuda a los modelos a aprovechar el conocimiento de diversas fuentes de texto, lo que resulta en un mejor rendimiento en el análisis de sentimientos.

Por esta razón, se utiliza RoBERTa (Robustly Optimized BERT Pretraining Approach) [96], ya que este modelo es una variante de BERT. Además, RoBERTa fue entrenado en un conjunto de datos mucho más grande con 160 GB de texto, lo cual es más de 10 veces mayor que el conjunto de datos utilizado para entrenar BERT. Por otro lado, RoBERTa ha utilizado un procedimiento de entrenamiento más efectivo al utilizar una técnica de enmascaramiento dinámico durante el entrenamiento que ayuda al modelo a aprender representaciones de palabras más robustas y generalizables.

5.3 Análisis de datos

El conjunto de datos Rest-Mex 2023 contiene 359,565 instancias de opiniones de turistas compartidas en TripAdvisor entre 2002 y 2022. Este conjunto de datos se divide en conjuntos de entrenamiento y prueba. El conjunto de entrenamiento contiene 251,702 instancias, lo que representa el 70% del conjunto de datos original, mientras que el conjunto de prueba contiene 107,863 instancias, que representa el 30% restante. Cada instancia del conjunto de entrenamiento tiene un título, una opinión que es el comentario emitido por el turista, la polaridad de la opinión (1, 2, 3, 4, 5), el país del lugar visitado (México, Cuba, Colombia) y el tipo de lugar del cual se emite la opinión (Hotel, Restaurante, Atractivo). Al mismo tiempo, cada instancia del conjunto de prueba tiene un identificador, un título y una opinión.

Se analizó y proceso el conjunto de entrenamiento del Rest-Mex 2023, el cual se dividió en conjuntos de entrenamiento y prueba. El conjunto de entrenamiento contiene 226,531 instancias, lo que representa el 90% del conjunto de entrenamiento de Rest-Mex 2023, mientras que el conjunto de prueba contiene 25,171 instancias que representan el 10% restante. Las Figuras 16, 17 y 18 presentan los tres tipos de distribución del conjunto de entrenamiento: polaridad, tipo y país. Se puede observar que este conjunto de datos está desbalanceado, especialmente en la distribución de polaridades. La clase mayoritaria representada por la etiqueta 5 de la polaridad tiene el 62.41% de las instancias, mientras que la clase minoritaria representada por la etiqueta 1 tiene solo el 2.29%. Además, los valores de polaridad negativa (1, 2) y el valor de polaridad neutral (3) tienen un número bajo de instancias en contraste con las polaridades positivas (valores 4 y 5). En la distribución por tipo, la clase mayoritaria es la etiqueta "Atractivo" con el 44.17% de las instancias, mientras que la clase minoritaria es la etiqueta "Restaurante" con el 25.61%. En la distribución por país, la clase mayoritaria es la etiqueta "México" con el 47.18% de las instancias, mientras que la clase minoritaria es la etiqueta "Cuba" con el 26.31%.

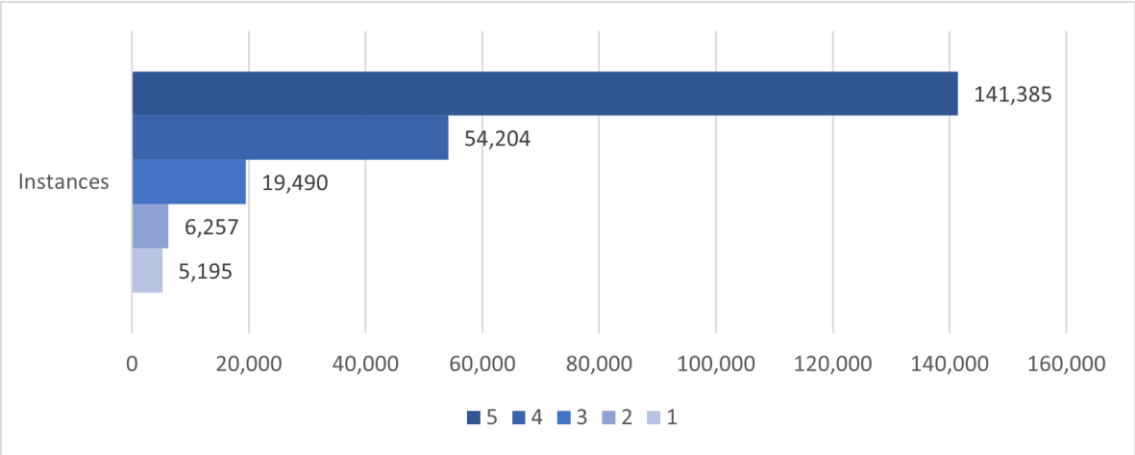

Figura 16. Distribución de la clase polaridad en el conjunto de entrenamiento.

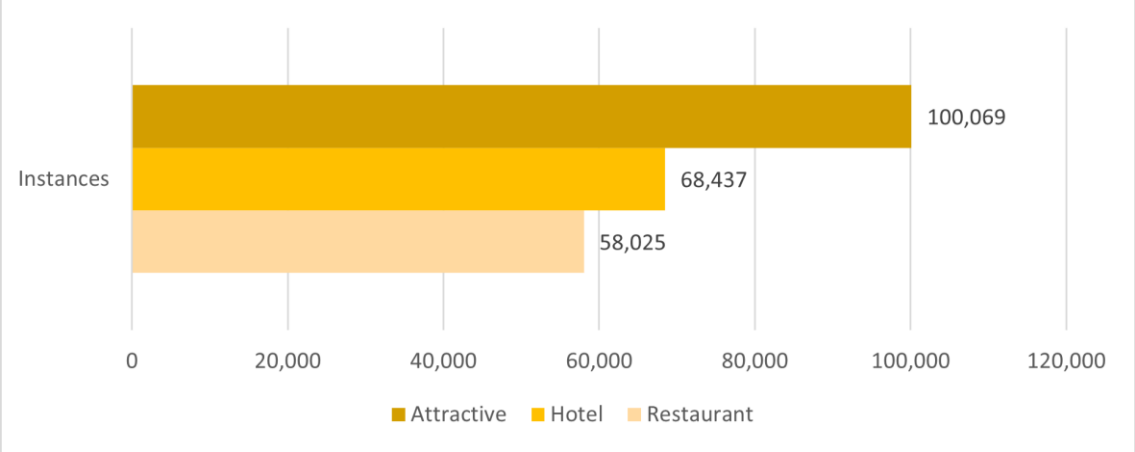

Figura 17. Distribución de la clase tipo en el conjunto de entrenamiento.

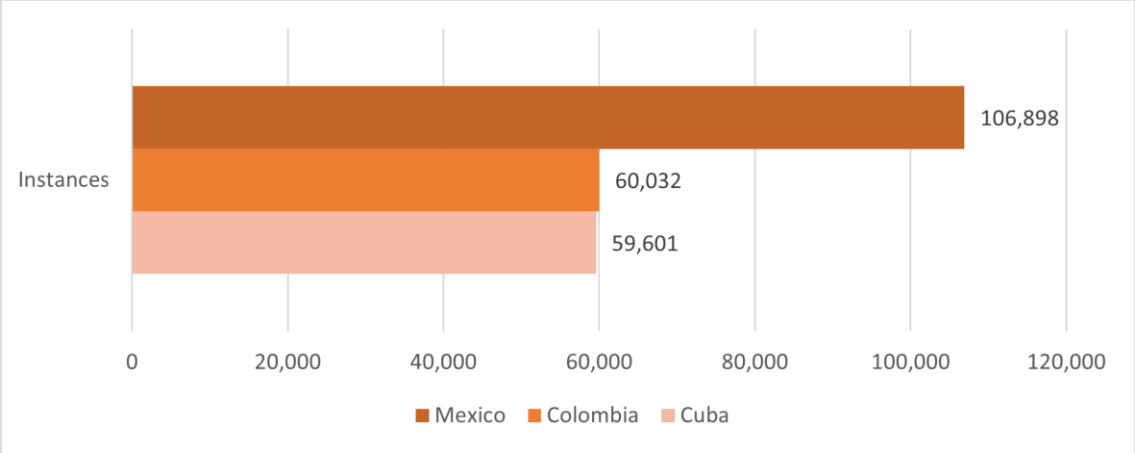

Figura 18. Distribución de la clase país en el conjunto de entrenamiento.

5.4 Metodología

Este apartado describe la metodología para abordar la tarea de análisis de sentimientos en Rest-Mex 2023. La Figura 19 presenta el diseño metodológico compuesto por 3 pasos principales: aumento de datos, adaptación de dominio y estrategias de entrenamiento. En el primer paso, utilizamos la técnica de traducción inversa para abordar el problema del desequilibrio de datos. En el segundo paso, realizamos un ajuste fino en un modelo de lenguaje preentrenado de RoBERTa para adaptarlo a un dominio turístico. En el tercer paso, diseñamos tres estrategias de entrenamiento para un modelo RoBERTa basado en el idioma español. Estos pasos se explican con más detalle a continuación.

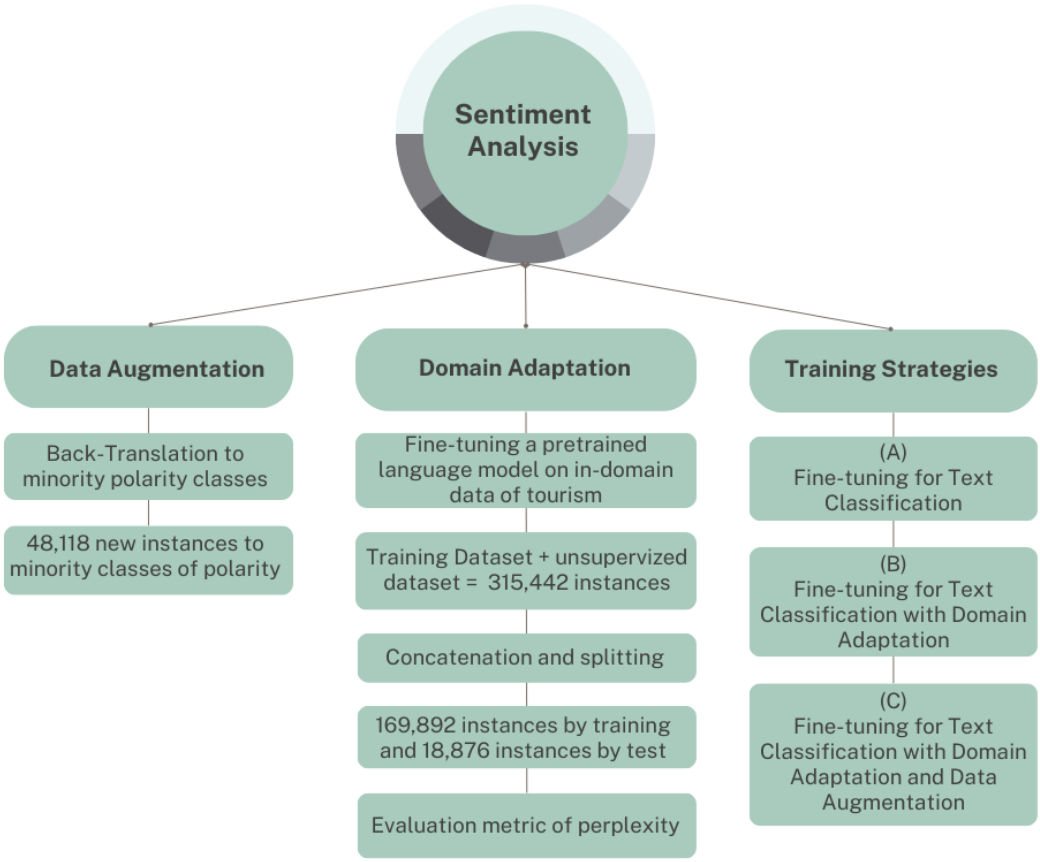

Figura 19. Diseño metodológico.

5.4.1 Aumento de datos

El desequilibrio de datos en el aprendizaje automático se refiere a una distribución desigual de las clases del conjunto de datos, y este problema es un desafío que necesita

más atención para resolverse [97]. Hay dos formas principales de abordar este problema: los métodos de submuestreo y los de sobremuestreo. Las técnicas de submuestreo disminuyen el número de instancias de la clase mayoritaria, mientras que los métodos de sobremuestreo aumentan el número de instancias de la clase minoritaria creando nuevos ejemplos o repitiendo algunos. La traducción inversa en inglés back-translation es una técnica de sobremuestreo que traduce un texto del idioma de destino al idioma fuente [98]. Existen varios modelos de transformers de código abierto para realizar traducciones automáticas de textos.

Utilizamos una técnica de sobremuestreo basada en la traducción inversa con transformer para generar nuevas reseñas por cada instancia con valores de polaridad negativa o neutral. Traducimos la reseña de texto de cada instancia del conjunto de datos de entrenamiento a un idioma de destino, y el resultado se traduce nuevamente a su idioma fuente, el español. Así, se emplean varios idiomas como idiomas de destino. Se generan tres nuevas instancias por cada instancia con etiqueta 1 utilizando los idiomas en inglés, francés y alemán como idiomas de destino. Se generaron dos nuevas instancias por cada instancia con etiqueta 2 utilizando los idiomas inglés y francés como idiomas de destino. Se desarrolló una nueva instancia por cada instancia con etiqueta 3 utilizando como idioma de destino el inglés. La Tabla 13 muestra los modelos de transformers de traducción utilizados para realizar la back-translation de clases minoritarias en el conjunto de datos de entrenamiento, el idioma fuente y el idioma de destino utilizados por estos modelos, y las etiquetas de las instancias traducidas por estos modelos.

Tabla 13. Modelos de traducción basado en transformer.

Model	**Source language**	**Target language**	**Label**
Helsinki-NLP/opus-mt-es-en	Spanish	English	1, 2, 3
facebook/nllb-200-distilled-600M	English	Spanish	1, 2, 3
Helsinki-NLP/opus-mt-es-fr	Spanish	French	1, 2
Helsinki-NLP/opus-mt-fr-es	French	Spanish	1, 2
Helsinki-NLP/opus-mt-es-de	Spanish	Deutsch	1
Helsinki-NLP/opus-mt-de-es	Deutsch	Spanish	1

La Figura 20 muestra la nueva distribución de la clase de polaridad en el conjunto de datos de entrenamiento, donde se incluyen 48,118 nuevas instancias asignadas a las etiquetas 1, 2 y 3. Generamos un conjunto de datos con estas nuevas instancias llamado analisis-sentimientos-textos-turisticos-mx-polaridadV3-DA.

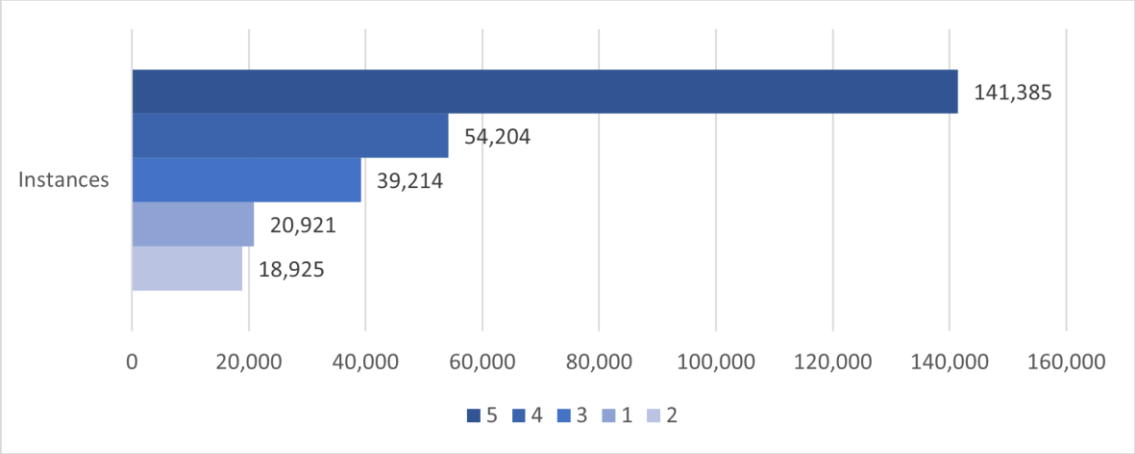

Figura 20. Distribución de la clase polaridad en el conjunto de entrenamiento con aumento de datos.

5.4.2 Adaptación de dominio

Los modelos de transformers se utilizan en muchas aplicaciones de PLN porque proporcionan buenos resultados con el aprendizaje por transferencia cuando el corpus utilizado para el preentrenamiento de un modelo no es demasiado diferente del corpus utilizado para el ajuste fino de la tarea específica en cuestión. Sin embargo, hay algunos casos en los que se desea ajustar finamente los modelos de lenguaje en sus datos antes de entrenar una cabeza específica para una tarea. Este proceso de ajuste fino de un modelo de lenguaje preentrenado en datos de dominio se llama adaptación de dominio [95]. Por ejemplo, un conjunto de datos de contratos legales dificulta el aprendizaje por transferencia con un modelo de transformer como BERT, porque este modelo procesará las palabras específicas del dominio de este conjunto de datos como tokens raros, y los resultados de rendimiento pueden no ser satisfactorios. Por esta razón, la adaptación de dominio es una alternativa para este problema, y puede aumentar el rendimiento de muchas tareas posteriores. Generalmente, se logra solo una vez [99].

Se ajusto finamente un modelo preentrenado llamado RoBERTa-base-bne [100], que ha sido entrenado con un gran corpus de español de 570 GB de texto limpio y deduplicado, obtenido a partir de un web crawling realizado por la Biblioteca Nacional de España desde 2009 hasta 2019. Usamos un conjunto de datos no supervisado con 63,740 instancias [87]. Se unió este conjunto de datos no supervisado con 251,702 instancias del conjunto de entrenamiento de Rest-Mex 2023 para producir 315,442

instancias de un nuevo conjunto de datos que contiene solo reseñas de texto sin etiquetas de polaridad, tipo y país. En la etapa de preprocesamiento, todos los textos de las instancias del conjunto de datos se tokenizan, concatenan y dividen en fragmentos de tamaño igual para evitar que los textos individuales se trunquen si su longitud es demasiado larga. Por lo tanto, este paso puede reducir la pérdida de información. El resultado del preprocesamiento es un conjunto de datos con 188,768 instancias tokenizadas con input_ids, attention_mask, word_ids y labels como características. En el ajuste fino, enmascaramos aleatoriamente el 15% de los tokens en cada lote de texto, como en el siguiente ejemplo.

- <s>Me ha parecido un museo excepcional,<mask> en una zona<mask> accesible de la Ciudad de México. Las salas son impresionantes carpintería, sobre todo las <mask> Teotihu<mask>án y la de la cultura Maya<mask> Existe la posibilidad de que te acompañe un especialista<mask> la visita y<mask><mask> pena. <mask>. . . </s>;

Como se observa, el token <mask> se ha insertado aleatoriamente en varias ubicaciones del texto y nuestro modelo debe predecir estos tokens durante el proceso de entrenamiento. Además, dividimos el conjunto de datos preprocesado en conjuntos de entrenamiento y prueba, donde el conjunto de entrenamiento contiene 169,892 instancias y el conjunto de prueba contiene 18,876 instancias. Los argumentos principales para el proceso de entrenamiento se muestran en la Tabla 14.

Tabla 14. Parámetros de entrenamiento para la adaptación de dominio.

Parameter	**Value**
batch_size	64
evaluation_strategy	epoch
number_of_epochs	3
learning_rate	2e-5
weight_decay	0.01
per_device_train_batch_size	batch_size
per_device_eval_batch_size	batch_size
fp16	True
logging_steps	length of training dataset // batch_size

La perplejidad [95] es una métrica estándar para evaluar el rendimiento de los modelos de lenguaje, ya que los modelos de lenguaje enmascarados no usan conjuntos de datos etiquetados. Esta métrica calcula las probabilidades que asigna a la siguiente palabra en todas las oraciones del conjunto de datos de prueba, donde las probabilidades altas significan que el modelo no está perplejo por las instancias de prueba. Utilizamos la exponencial de la pérdida de entropía cruzada como una definición matemática de perplejidad. Se calculó la perplejidad evaluando la función de transformers, donde un puntaje de perplejidad más bajo significa un mejor modelo de lenguaje. La perplejidad del modelo sin ejecutar el entrenamiento es 28.03. En contraste, la perplejidad de nuestro modelo entrenado es 6.05, lo cual es más bajo que el primer resultado, representando buenos resultados para la adaptación de dominio en RoBERTa-base-bne. El resultado del modelo de lenguaje enmascarado se llama roberta-base-bne-finetuned-TripAdvisorDomainAdaptation.

5.4.3 Estrategias de entrenamiento

Las tres estrategias principales para el entrenamiento se diseñan utilizando un modelo de lenguaje enmascarado llamado RoBERTabase-bne [100], que ha sido preentrenado utilizando un gran corpus en español con 570 GB de texto limpio y deduplicado. Además, los textos fueron obtenidos mediante un rastreo web realizado por la Biblioteca Nacional de España entre 2009 y 2019. A continuación, explicamos las estrategias.

En la primera estrategia, ajustamos finamente RoBERTa-base-bne para clasificar la polaridad, tipo y país en textos de reseñas de TripAdvisor. Utilizamos el conjunto de entrenamiento Rest-Mex 2023 y lo dividimos de manera estratificada por cada clase en conjuntos de entrenamiento y prueba. El conjunto de entrenamiento contiene el 70% de las instancias del conjunto de entrenamiento Rest-Mex 2023, mientras que el conjunto de prueba contiene el 30% restante. Además, se combinó el título y la reseña de las instancias como la única característica, se renombró la clase objetivo como etiquetas y eliminamos el resto de las características. El resultado de este proceso son tres conjuntos de datos llamados analisis-sentimiento-textos-turisticos-mx-polaridad, analisis-sentimiento-textos-turisticos-mxtipo y analisis-sentimiento-textos-turisticos-mx-pais (ver en https://huggingface.co/alexcom/).

Luego, para la clasificación con transformers, modificamos las etiquetas de polaridad de [1, 2, 3, 4, 5] a [0, 1, 2, 3, 4], las etiquetas de tipo de [Attractive, Hotel, Restaurant] a [0, 1, 2] y las etiquetas de país de [Mexico, Colombia, Cuba] a [0, 1, 2]. Se tokenizó cada conjunto de datos con el AutoTokenizer de RoBERTa-base-bne de transformers para generar las características de etiquetas, input_ids y attention_mask. Además, se seleccionó el número de etiquetas para cada clase. Por ejemplo, el número de etiquetas

para la clasificación de polaridad es 5 y el número de etiquetas para la clasificación de tipo y país es 3.

Se utilizó la métrica F1 para la evaluación del modelo y mostramos algunos parámetros de entrenamiento para la clasificación de texto en la Tabla 15. Además, utilizamos solo 20,000 instancias del conjunto de entrenamiento. Con este proceso, generamos tres modelos llamados roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticos-mx-polaridad para la clasificación de polaridad, roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticos-mxtipo para la clasificación de tipo y roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticosmx-pais (ver en https://huggingface.co/vg055/) para la clasificación de país.

Tabla 15. Parámetros de entrenamiento para la clasificación de texto.

Parameter	**Value**
batch_size	16
num_train_epochs	2
evaluation_strategy	epoch
learning_rate	2e-5
weight_decay	0.01
per_device_train_batch_size	batch_size
per_device_eval_batch_size	batch_size
logging_steps	length of training dataset // (2*batch_size*num_train_epochs)

En la segunda estrategia, ajustamos finamente roberta-base-bne-finetunedTripAdvisorDomainAdaptation para clasificar la polaridad, tipo y país en textos de reseñas de TripAdvisor. Utilizamos el conjunto de entrenamiento Rest-Mex 2023 y lo dividimos de manera estratificada por cada clase en conjuntos de entrenamiento y prueba. El conjunto de entrenamiento contiene el 90% de las instancias del conjunto de entrenamiento Rest-Mex 2023, mientras que el conjunto de prueba contiene el 10% restante. Procesamos el conjunto de datos de la misma manera que en la primera estrategia y generamos tres conjuntos de datos llamados analisis-sentimientos-textos-turisticos-mx-polaridadV2, analisis-sentimientos-textos-turisticos-mx-tipoV2 y analisis-sentimientos-textos-turisticos-mx-paisV2. Luego, realizamos los mismos pasos de la primera estrategia en la tokenización, preprocesamiento y

entrenamiento, pero ahora con el modelo adaptado al dominio del turismo y utilizando todas las instancias del conjunto de entrenamiento para el entrenamiento. Con este proceso, generamos tres modelos llamados roberta-base-bne-finetunedTripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-polaridad para la clasificación de polaridad, roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-tipo para la clasificación de tipo y roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-pais para la clasificación de país.

En la tercera estrategia, ajustamos finamente roberta-base-bne-finetuned-TripAdvisorDomainAdaptation para clasificar la polaridad en textos de reseñas de TripAdvisor. A diferencia de las otras estrategias, se utilizó el conjunto de datos analisis-sentimientos-textos-turisticos-mx-polaridadV3-DA, que se basa en el conjunto de entrenamiento Rest-Mex 2023, pero con instancias de las clases negativa y neutral aumentadas mediante una técnica de aumento de datos aplicada previamente. Se repitió el proceso de la segunda estrategia con este conjunto de datos y se generó un modelo llamado roberta-base-bne-finetunedTripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-polaridadDA-V3 para la clasificación de polaridad.

Se realizaron en total 23 experimentos: 19 experimentos para la identificación de polaridad, 2 experimentos para la clasificación de tipo y 2 experimentos para la identificación de país. Se dirigió la investigación sobre la identificación de polaridad debido a que el problema es más complejo por varias razones, como el desequilibrio de datos y la clasificación de 5 clases. Se incluyeron solo los resultados más representativos de estos experimentos para cada estrategia de entrenamiento en la siguiente sección.

5.5 Resultados

Los resultados obtenidos durante el período de desarrollo se presentan en la Figura 21, donde la segunda estrategia muestra el mejor resultado, seguida por la tercera y la primera estrategia. La clasificación de polaridad, tipo y país relacionada con la primera estrategia mejora con la adaptación de dominio aplicada en la segunda estrategia. Sin embargo, la técnica de aumento de datos utilizada en la tercera estrategia no representa una mejora respecto a la segunda estrategia, lo que subraya la importancia de revisar detalladamente las técnicas de aumento de datos para trabajos futuros.

Se generaron tres salidas basadas en estas estrategias previas y en el conjunto de prueba Rest-Mex 2023. Todas las salidas utilizan la segunda estrategia para la clasificación de tipo y país. La salida 1 utiliza la primera estrategia para la clasificación de polaridad, la salida 2 emplea la segunda estrategia para la clasificación de polaridad, y la salida 3 aplica la tercera estrategia para clasificar la polaridad. Enviamos estas salidas a Rest-

Mex 2023, y los resultados obtenidos en el proceso de prueba de Rest-Mex 2023 se presentan en la Figura 22, donde el mejor resultado se logró con la salida 2, seguido por la salida 1 y la salida 3. Sin embargo, la diferencia entre las tres salidas es muy baja debido a que solo la clasificación de polaridad las diferencia. Por otro lado, observamos que la adaptación de dominio es una excelente alternativa para mejorar los resultados. No obstante, es esencial buscar otros métodos para abordar el problema del desequilibrio de clases en la clasificación de polaridad.

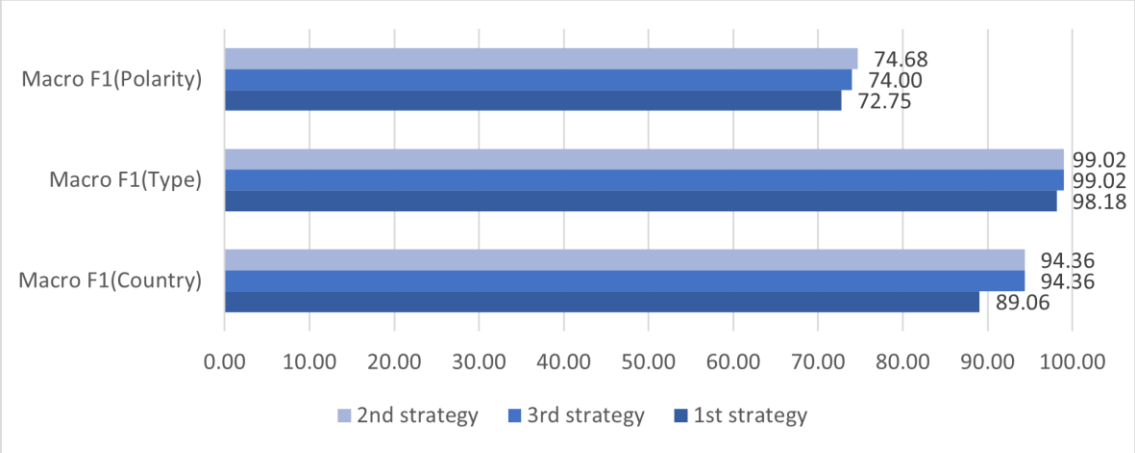

Figura 21. Resultados de las estrategias de entrenamiento.

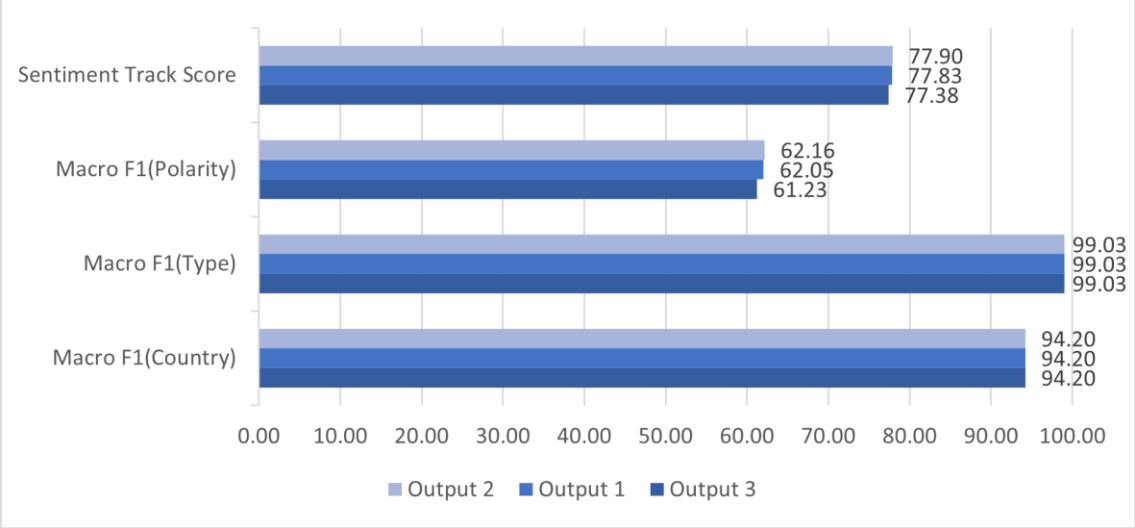

Figura 22. Resultados del envío de salidas para el Rest-Mex 2023.

Se logró el primer lugar en el ranking con la salida 2 en la tarea de análisis de sentimientos de RestMex 2023, en la que participaron 18 equipos de todo el mundo. Además, las tres salidas ocuparon los tres primeros lugares en esta tarea compartida.

5.5 Conclusiones

El equipo LKE-IIMAS, logró los mejores resultados en la tarea de análisis de sentimientos en Rest-Mex 2023 con una adaptación de dominio a un modelo preentrenado basado en RoBERTa, el cuál fue entrenado con un gran conjunto de datos en español. Se desarrollaron tres pasos principales en nuestra metodología: aumento de datos, adaptación de dominio y estrategias de entrenamiento. Además, generamos varios conjuntos de datos con diferentes tipos de estratificación clasificando la polaridad, el tipo y el país. Ajustamos finamente un modelo de lenguaje enmascarado para aplicar la adaptación de dominio y evaluamos su perplejidad. Así mismo, ajustamos un modelo de lenguaje para la clasificación de textos por polaridad, tipo y país. Evaluamos el resultado de estos modelos con la métrica F1 para seleccionar las tres mejores estrategias y generar tres envíos de salida a Rest-Mex 2023. La adaptación de dominio mejoró los resultados. Sin embargo, es importante analizar otros métodos de aumento de datos que produzcan nuevas instancias con calidad.

Agradecimientos

Este trabajo se ha llevado a cabo con el apoyo del proyecto DGAPA-UNAM PAPIIT número TA101722. Los autores también agradecen al Laboratorio de Ingeniería del Lenguaje y del Conocimiento (LKE) de la Benemérita Universidad Autónoma de Puebla, al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM) y al CONACYT por los recursos computacionales proporcionados a través de la Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje del Laboratorio de Supercómputo del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE). También queremos agradecer al Ing. Román Osorio por apoyar la administración estudiantil del proyecto.

Capítulo 6: Un sistema de recomendación multidominio basado en Transformer para el e-commerce

Los sistemas de recomendación (RS) se erigen como una de las aplicaciones más vitales de la inteligencia artificial (IA), la ciencia de datos y las técnicas avanzadas de análisis. Su integración en nuestra vida diaria abarca desde el trabajo, las operaciones comerciales, el estudio, el entretenimiento y la socialización [101]. Además, el sistema de recomendación sirve como una herramienta potente para tomar decisiones bien informadas, efectivas y eficientes en una amplia gama de ítems, incluidos productos comerciales, películas, listas de reproducción, puntos de interés, hoteles, restaurantes, amigos, viajes, carreras y más [102]. Grandes empresas internacionales como Amazon, Spotify y Netflix han integrado estos sistemas en sus servicios principales. Al abordar la sobrecarga de información, mejorar las experiencias de usuario y reducir las tasas de abandono, los RS impactan significativamente en los beneficios de las empresas.

Tradicionalmente, las técnicas en los sistemas de recomendación han girado en torno al filtrado de contenido, que calcula elementos similares basados en sus metadatos, el filtrado colaborativo, que identifica usuarios similares utilizando un conjunto de calificaciones en inglés rankings, y enfoques híbridos que combinan ambas técnicas para optimizar el rendimiento de los RS. La investigación de vanguardia ha identificado problemas principales como el problema de arranque en frío, la escalabilidad y la falta de novedad en las recomendaciones, todos los cuales son abordados por los enfoques híbridos [31]. Sin embargo, estas técnicas suelen basarse en todas las interacciones históricas entre usuario e ítem, incluidos clics, compras, vistas, etc., para aprender las preferencias a largo plazo y estáticas del usuario sobre los ítems, haciendo la suposición subyacente de que todas las interacciones históricas son igualmente importantes para las preferencias actuales del usuario [102]. Esta suposición pasa por alto dos puntos clave: en primer lugar, la importancia de las preferencias recientes a corto plazo y el contexto sensible al tiempo en las elecciones del usuario, y en segundo lugar, la naturaleza dinámica de las preferencias del usuario, que evolucionan con el tiempo.

Además, se observa que solo un pequeño número de interacciones históricas representan las interacciones más recientes de los usuarios, que contienen las preferencias recientes a corto plazo de los usuarios [103].

En los últimos años, ha surgido un nuevo paradigma de sistemas de recomendación, conocido como sistemas de recomendación basados en sesiones (SBRS), para abordar las preferencias a corto plazo y dinámicas de los usuarios, con el objetivo de generar recomendaciones más oportunas y precisas [69]. Este paradigma es una subárea de los sistemas de recomendación secuenciales (SRS), lo que hace que SBRS esté estrechamente relacionado con SRS en cuanto a entrada, salida y mecanismo de recomendación. El objetivo principal de SBRS es aprender las dependencias incrustadas en secuencias o sesiones para inferir las preferencias dinámicas de los usuarios. En un SBRS, las preferencias de los usuarios se aprenden a partir de sesiones, cada una compuesta por múltiples interacciones entre usuario e ítem que ocurren en un período continuo de tiempo. Al considerar cada sesión como la unidad de entrada fundamental, SBRS puede capturar tanto las preferencias a corto plazo de un usuario a partir de sus sesiones recientes como la dinámica de preferencias, reflejando los cambios en las preferencias de una sesión a otra, lo que permite recomendaciones más precisas y oportunas.

Los avances recientes en los sistemas de recomendación secuenciales y basados en sesiones han sido impulsados por mejoras en la arquitectura de modelos del lenguaje y técnicas de preentrenamiento originadas en el campo del Procesamiento del Lenguaje Natural (NLP), particularmente las arquitecturas Transformer. Estas arquitecturas han suplantado a las redes neuronales convolucionales y recurrentes en tareas de modelado del lenguaje debido a su entrenamiento paralelo eficiente, escalabilidad con datos de entrenamiento y tamaño del modelo, y efectividad en la modelización de secuencias de largo alcance. Además, los Transformers han facilitado el desarrollo de modelos de mayor capacidad y han introducido técnicas de aumento de datos y entrenamiento que mejoran significativamente la eficacia de los sistemas de recomendación secuenciales.

El procesamiento secuencial de las interacciones de usuario en las recomendaciones secuenciales y basadas en sesiones es similar con la tarea del modelado del lenguaje (LM), lo que ha llevado a la adaptación de muchas arquitecturas de Transformer de NLP, como la biblioteca Transformers4Rec [104]. Esta biblioteca de código abierto, construida sobre la biblioteca Transformers de HuggingFace, comparte el objetivo de extender los avances de los Transformers basados en NLP a la comunidad de sistemas de recomendación, haciendo que estos avances sean fácilmente accesibles para tareas de recomendación secuenciales y basadas en sesiones.

Por lo tanto, este estudio construye un sistema de recomendación multidominio basado en Transformer para el comercio electrónico utilizando los últimos avances en NLP como Transformers4Rec, facilitando un análisis empírico de la recomendación basada

en sesiones en datos específicos de un único dominio o datos específicos de múltiples dominios obtenidos de un conjunto de datos de reseñas de Amazon [105]. Se comparan varias técnicas de entrenamiento, incluyendo el Modelo de Lenguaje Causal (CLM), el Modelo de Lenguaje Enmascarado (MLM), el Modelo de Lenguaje de Permutación (PLM) y la Detección de Token de Reemplazo (RTD) por una arquitectura Transformer conocida como XLNet [106], dentro de una Unidad Recurrente Cerrada (GRU).

Este trabajo está estructurado de la siguiente manera: La subsección 6.1 introduce el contexto sobre los sistemas de recomendación basados en sesiones y los Transformers. La subsección 6.2 discute trabajos relacionados. La subsección 6.3 describe la metodología propuesta. La subsección 6.4 presenta los resultados, y la subsección 6.5 concluye el estudio.

6.1 Fundamentos

La sección de fundamentos proporciona una visión general de los sistemas de recomendación basados en sesiones y del modelado del lenguaje basado en Transformers, destacando sus fundamentos y su importancia en el análisis de diversas técnicas de entrenamiento dentro de estas arquitecturas.

6.1.1 Sistema de recomendación basado en sesiones

Los sistemas de recomendación basados en sesiones se centran en comprender y predecir las preferencias del usuario basadas en interacciones secuenciales o sesiones. Estos sistemas tienen como objetivo capturar la naturaleza dinámica del comportamiento del usuario al considerar la secuencia de acciones tomadas dentro de una sesión. Al aprovechar estos datos secuenciales, los sistemas de recomendación basados en sesiones pueden ofrecer recomendaciones personalizadas que se alinean con las preferencias e intereses inmediatos de los usuarios.

La diferencia entre los datos de sesión y los datos de secuencia radica en su estructura y organización. Una sesión se refiere a una lista finita de interacciones, que pueden estar ordenadas o desordenadas. Cuando las interacciones dentro de una sesión están organizadas cronológicamente, se denomina sesión ordenada. Por el contrario, si las interacciones no están organizadas cronológicamente, se denomina sesión desordenada. Varias sesiones pueden ocurrir en diferentes momentos, formando colectivamente los datos de sesión de un usuario. Estas sesiones están delimitadas por diversos límites, con intervalos de tiempo potencialmente no idénticos entre ellas. Por ejemplo, considera la Figura 23, que ilustra un ejemplo de datos de sesión de usuario. En este escenario, se representan tres sesiones, cada una separada por límites que ocurren en

intervalos de 2 semanas y 4 semanas a lo largo del tiempo. La sesión 1 comprende tres interacciones entre usuario e ítem, mientras que las sesiones 2 y 3 contienen cada una dos interacciones entre usuario e ítem. A pesar de las variaciones en el número de interacciones, todas las sesiones contribuyen a los datos de sesión del usuario en general, proporcionando información sobre su comportamiento y preferencias a lo largo del tiempo.

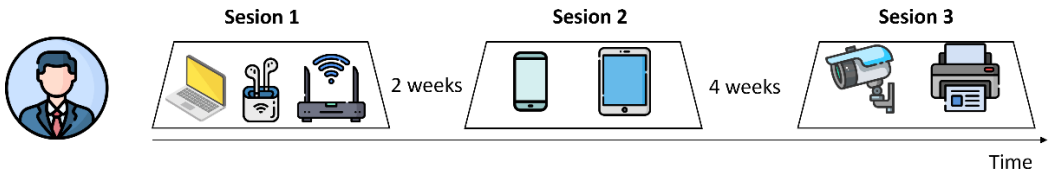

Figura 23. Este es un ejemplo de datos de sesión del usuario que contiene tres sesiones divididas por límites de 2 semanas y 4 semanas a lo largo del tiempo.

Una secuencia se refiere a una lista de elementos (ítems) históricos organizados en un orden específico, como una secuencia de identificadores de ítems. En el contexto de los datos del usuario, una secuencia representa típicamente una serie de interacciones o eventos, donde cada elemento en la secuencia indica una acción específica tomada por el usuario a lo largo del tiempo. A diferencia de las sesiones, las secuencias no tienen múltiples límites y se caracterizan por una única secuencia continua de eventos. Por ejemplo, considera la Figura 24, que ilustra un ejemplo de datos de secuencia del usuario. En esta representación, los datos de secuencia del usuario consisten en una sola secuencia que comprende cinco películas vistas en orden cronológico. Cada película en la secuencia representa un evento o interacción distinta, formando una representación clara y ordenada del historial de visualización del usuario. A diferencia de las sesiones, que pueden contener múltiples sesiones separadas por límites, una secuencia encapsula todas las interacciones dentro de una única secuencia continua, proporcionando una visión general completa del comportamiento del usuario.

Figura 24. Este es un ejemplo de datos de secuencia del usuario que contiene una secuencia de cinco películas vistas en orden cronológico por el usuario.

Efectivamente, los sistemas que utilizan datos de sesión no ordenados suelen basarse en dependencias basadas en la co-ocurrencia. En este contexto, la co-ocurrencia se refiere a la aparición de ítems juntos dentro de la misma sesión, independientemente de su orden. Estos sistemas analizan patrones de co-ocurrencia de ítems para identificar

asociaciones y realizar recomendaciones basadas en ítems observados con frecuencia juntos en sesiones.

Por otro lado, los sistemas que utilizan datos de sesión ordenados o secuenciales aprovechan las dependencias secuenciales. Aquí, el orden de las interacciones dentro de las sesiones es crucial, ya que refleja la secuencia temporal de las acciones del usuario. Al considerar el orden secuencial de las interacciones, estos sistemas capturan las dependencias entre ítems basadas en la secuencia en la que fueron encontrados. Este enfoque permite modelar las preferencias y el comportamiento del usuario a lo largo del tiempo, lo que permite generar recomendaciones personalizadas y contextualmente relevantes.

El objetivo de un sistema de recomendación basado en sesiones es predecir la parte desconocida de una sesión, que podría ser un ítem o un conjunto de ítems, o incluso la sesión futura, como la próxima cesta de compra. Esta tarea está definida por cinco entidades clave:

- Usuarios, junto con sus propiedades.
- Ítems, junto con sus propiedades.
- Acciones, que abarcan las interacciones de los usuarios con los elementos, como clics, vistas y compras.
- Interacciones, representadas como tripletes de [usuario, acción, ítem], que capturan la relación entre usuarios, acciones y ítems.
- Sesiones, cada una caracterizada por propiedades como la duración de la sesión, el orden interno de las interacciones, los tipos de acciones, la información del usuario y la estructura de datos.

Existen varias aproximaciones líderes a los sistemas de recomendación basados en sesiones, incluyendo métodos convencionales, técnicas de representación latente y enfoques de redes neuronales profundas. Estas aproximaciones se discuten en detalle en la literatura, proporcionando información sobre sus fortalezas, debilidades y aplicaciones para abordar los desafíos de la recomendación basada en sesiones [102] [69].

Enfoques convencionales

Para capturar las dependencias inherentes en los datos de sesión para recomendaciones basadas en sesiones, se emplean técnicas de minería de datos o aprendizaje automático. Las siguientes técnicas son comúnmente utilizadas para este propósito:

- Minería de patrones/reglas.
- Vecinos más cercanos (K nearest neighbour).
- Cadena de Markov.
- Modelo probabilístico generativo.

Enfoques basados en representación latente

Se generan representaciones latentes de baja dimensionalidad para cada interacción dentro de las sesiones utilizando modelos superficiales. Estas representaciones aprendidas codifican dependencias informativas entre las interacciones, facilitando la generación de recomendaciones basadas en sesiones posteriores. Los siguientes métodos son comúnmente empleados para este propósito:

- Modelo de factores latentes.
- Representación distribuida.

Enfoques basados en redes neuronales profundas

Estos enfoques aprovechan la capacidad para modelar dependencias complejas intra e inter-sesión para recomendaciones. La clasificación de estas técnicas se divide en: (1) Redes neuronales profundas básicas, donde se utiliza una arquitectura fundamental de redes neuronales como las Redes Neuronales Recurrentes (RNN). (2) Modelos avanzados, que emplean mecanismos o modelos sofisticados como modelos de atención.

- Redes neuronales profundas básicas:
  - Redes neuronales recurrentes.
  - Redes neuronales perceptrón multicapa.
  - Redes neuronales convolucionales.
  - Redes neuronales de grafos.
- Modelos avanzados:
  - Modelo de atención.
  - Redes neuronales con memoria.
  - Modelo de mezcla.
  - Modelo generativo.
  - Aprendizaje por refuerzo.

6.1.2 Transformers

En este estudio, nos enfocamos en modelos avanzados, específicamente modelos de atención conocidos como Transformers, los cuales han ganado una popularidad significativa en la comunidad de Procesamiento del Lenguaje Natural (PLN). Propuestos por Vaswani et al. en 2017 [92], esta arquitectura neural ha demostrado un rendimiento notable en diversas tareas de PLN. En el núcleo de los Transformers se encuentra un mecanismo de autoatención, que enfatiza las relaciones contextuales entre palabras o tokens en una secuencia. Una de las principales ventajas de los Transformers

es su capacidad para manejar eficazmente dependencias a largo plazo. Esto se logra a través del mecanismo de autoatención, que permite al modelo evaluar la importancia de diferentes palabras o tokens en la secuencia mientras procesa la secuencia completa simultáneamente. Este mecanismo, también conocido como atención de producto punto escalado, permite al modelo calcular los pesos de atención para cada palabra/token en función de sus relaciones con otras palabras/tokens en la secuencia de entrada. Al aprovechar la atención, el modelo puede centrarse en partes relevantes de la entrada durante las etapas de codificación y decodificación, mejorando su capacidad para capturar dependencias intrincadas y producir predicciones precisas.

Estas arquitecturas se han categorizado según sus tareas de preentrenamiento, que son esenciales para aprender representaciones lingüísticas universales. Hay tres tipos principales de enfoques de aprendizaje utilizados.

- Aprendizaje supervisado: En este enfoque, el modelo aprende a partir de datos etiquetados, donde cada entrada está asociada con una etiqueta de salida correspondiente. El aprendizaje supervisado se utiliza comúnmente en tareas donde la verdad fundamental está disponible durante el entrenamiento, lo que permite que el modelo aprenda a predecir la salida correcta.
- Aprendizaje no supervisado: El aprendizaje no supervisado implica entrenar el modelo en datos no etiquetados, donde el objetivo es aprender patrones y estructuras de los datos sin orientación explícita o etiquetas. Este enfoque se utiliza a menudo para descubrir patrones o representaciones subyacentes en los datos sin la necesidad de ejemplos etiquetados.
- Aprendizaje auto-supervisado: El aprendizaje auto-supervisado se encuentra dentro de la categoría más amplia de aprendizaje no supervisado, pero implica la creación de tareas similares a las supervisadas a partir de los propios datos de entrada. En lugar de depender de etiquetas externas, el modelo genera sus propias señales de supervisión a partir de los datos de entrada. Este enfoque ha ganado popularidad debido a su capacidad para aprovechar de manera efectiva grandes cantidades de datos no etiquetados.

Las tareas de preentrenamiento suelen utilizar el aprendizaje auto-supervisado en los transformers, y estas tareas se presentan a continuación.

Modelado de Lenguaje Enmascarado (MLM)

Varias tareas de PLN han logrado un éxito notable al preentrenar codificadores de texto para aprender de contextos bidireccionales. Uno de los enfoques de preentrenamiento más prominentes es el Modelado de Lenguaje Enmascarado, conocido por su simplicidad conceptual y efectividad empírica [107]. Este enfoque implica enmascarar

una parte de los tokens de las oraciones de entrada y luego entrenar al modelo para predecir los tokens enmascarados basándose en el contexto circundante [108].

Modelado de Lenguaje Causal (CLM)

El Modelado de Lenguaje Causal está diseñado para predecir el próximo elemento en una secuencia de manera autorregresiva, lo que lo convierte en una de las aplicaciones fundamentales del modelo Transformer [109]. Este enfoque de preentrenamiento se emplea comúnmente en tareas de generación de texto, donde el modelo considera solo el contexto pasado y no el futuro al realizar predicciones, como se demuestra en modelos como GPT-2.

Modelado de Lenguaje de Permutación (PLM)

El Modelado de Lenguaje de Permutación es otra técnica de preentrenamiento utilizada en modelos Transformer. En este enfoque, se permutan aleatoriamente las palabras en las secuencias de entrada y luego el modelo se entrena para predecir la posición original de cada palabra permutada en la secuencia. Este método fomenta al modelo a capturar relaciones semánticas más profundas y a comprender el contexto global de la oración, en lugar de simplemente depender de las relaciones locales entre las palabras [108].

Detección de Tokens Reemplazados (RTD)

La Detección de Tokens Reemplazados es una técnica de preentrenamiento discriminatoria que implica entrenar un generador para producir tokens reemplazados y un discriminador para diferenciar entre tokens reales y reemplazados. Este método mejora la eficiencia del preentrenamiento al reducir la sobrecarga computacional en el encabezado en comparación con enfoques anteriores de modelado de lenguaje enmascarado [110].

6.2 Trabajos relacionados

El estudio titulado *BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer* [111] presentó un modelo de recomendación secuencial que aprovecha la atención automática bidireccional profunda para modelar secuencias de comportamiento del usuario. Los autores emplearon el objetivo Cloze, comúnmente utilizado en tareas de comprensión del lenguaje, para la recomendación secuencial. Este objetivo implica predecir elementos enmascarados aleatoriamente en la secuencia considerando conjuntamente su contexto izquierdo y derecho. Además, los autores realizaron experimentos en cuatro conjuntos de datos y demostraron que su modelo superaba a otros modelos secuenciales en términos de rendimiento de recomendación.

El estudio titulado *Exploiting deep transformer models in textual review-based recommender systems* [112] destaca que las reseñas textuales contienen información pertinente que puede inferir efectivamente las preferencias del usuario sobre los ítems. El estudio resalta que los modelos de aprendizaje profundo capturan mejor las interacciones entre usuario e ítem a partir de reseñas textuales en comparación con enfoques de recomendación tradicionales, mejorando así el rendimiento predictivo. Los autores emplearon y analizaron modelos profundos de transformers para sistemas de recomendación basados en reseñas, observando que estos modelos pueden extraer representaciones de usuario e ítem interpretables y relevantes de manera más efectiva que las redes de aprendizaje profundo tradicionales. Los resultados demuestran que el modelo de transformer con mejor rendimiento logró una mejora relativa máxima en el RMSE del 4.6% y un MAE del 7.4% con productos electrónicos de Amazon en comparación con el mejor resultado de las redes de aprendizaje profundo tradicionales.

El estudio titulado *Transformers4Rec: Bridging the Gap between NLP and Sequential/Session-Based Recommendation* [104] abordó la creciente disparidad entre los campos de Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) y la recomendación secuencial/basada en sesiones. Los autores desarrollaron Transformers4Rec, una biblioteca de código abierto destinada a cerrar esta brecha al proporcionar varias arquitecturas de transformers adaptadas para recomendaciones secuenciales y basadas en sesiones. Lograron resultados prometedores, especialmente en la predicción del próximo clic para sesiones de usuarios, a pesar de que las longitudes de secuencia son mucho más cortas que las comúnmente encontradas en tareas de NLP. Además, sus experimentos demostraron que las arquitecturas superiores produjeron un mejor rendimiento en dos conjuntos de datos de comercio electrónico, mientras mantenían un rendimiento similar a los modelos de referencia en dos conjuntos de datos de noticias.

6.3 Análisis de datos

Se utilizó un conjunto de datos de reseñas de Amazon obtenido de [105], que consta de 233.1 millones de reseñas recopiladas desde mayo de 1996 hasta octubre de 2018. Este conjunto de datos ofrece varios subconjuntos, que incluyen datos de reseñas completas, datos solo de calificaciones, datos con 5 calificaciones, datos por categoría y subconjuntos más pequeños diseñados para experimentación, como k-calificaciones y solo calificaciones. A partir del subconjunto de datos por categoría de datos de reseñas completas, seleccionamos 15 subconjuntos para nuestro análisis. La Tabla 16 muestra los detalles de estos subconjuntos, que en conjunto contienen 102,395,764 reseñas y 5,993,235 entradas de metadatos para productos.

Tabla 16. Datos por categoría de productos de Amazon.

No	Category	Reviews	Products
1	All_Beauty	371,345	32,992
2	AMAZON_FASHION	883,636	186,637
3	Musical_Instruments	1,512,530	120,400
4	Industrial_and_Scientific	1,758,333	167,524
5	Video_Games	2,565,349	84,893
6	Grocery_and_Gourmet_Food	5,074,160	287,209
7	Patio_Lawn_and_Garden	5,236,058	279,697
8	Office_Products	5,581,313	315,644
9	Pet_Supplies	6,542,483	206,141
10	Toys_and_Games	8,201,231	634,414
11	Movies_and_TV	8,765,568	203,970
12	Cell_Phones_and_Accessories	10,063,255	590,269
13	Sports_and_Outdoors	12,980,837	962,876
14	Electronics	20,994,353	786,868
15	Home_and_Kitchen	21,928,568	1,301,225
	Total	102,395,764	5,993,235

Los datos de las reseñas contienen los siguientes campos:

- reviewerID: ID del revisor.
- asin: ID del producto.
- reviewerName: nombre del revisor.
- vote: votos útiles de la reseña.
- verify: si la reseña está verificada.
- style: un diccionario de metadatos del producto.
- reviewText: texto de la reseña.
- overall: calificación del producto.
- summary: resumen de la reseña.
- unixReviewTime: tiempo de la reseña (tiempo UNIX).
- reviewTime: tiempo de la reseña (en formato sin procesar).
- image: imágenes que los usuarios publican después de haber recibido el producto.

Los metadatos de los productos contienen los siguientes campos:

- asin: ID del producto.
- title: nombre del producto.
- feature: características del producto en formato de viñetas.
- description: descripción del producto.
- price: precio en dólares estadounidenses (en el momento de la extracción de datos).
- imageURL: URL de la imagen del producto.

- related: productos relacionados (también comprados, también vistos, comprados juntos, comprar después de ver).
- salesRank: información sobre el rango de ventas.
- brand: nombre de la marca.
- categories: lista de categorías a las que pertenece el producto.
- tech1: la primera tabla de detalles técnicos del producto.
- tech2: la segunda tabla de detalles técnicos del producto.
- similar: tabla de productos similares.

El preprocesamiento de datos es crucial para adaptar el conjunto de datos a las tareas de recomendación basada en sesiones, dado que contiene información relacionada con técnicas de filtrado basado en contenido, filtrado colaborativo y enfoques híbridos, que son técnicas de recomendación tradicionales. El objetivo es incorporar características adicionales que complementen los identificadores de productos en las sesiones, como la calificación general, el estado de verificación, el precio, la categoría, la marca, etc. Además, es necesario agrupar las revisiones por fechas para crear sesiones de usuario.

En el primer paso del preprocesamiento, los subconjuntos se procesan para eliminar instancias duplicadas y eliminar columnas que no son relevantes para los datos de revisiones y productos, incluyendo *reviewerName, reviewText, summary, vote, style, image, title, feature, date, imageURL, imageURLHighRes, description, also_view, also_buy, fit, details, similar_item, tech1* y *tech2*. Además, concatenamos *reviewerID* y *unixReviewTime* para formar el ID de la sesión, lo que indica que una sesión comprende todas las compras realizadas por un usuario en la misma fecha.

Durante este paso inicial, calculamos y guardamos el número de sesiones por fecha y el número de categorías asociadas con cada fecha. Luego, seleccionamos una semana con el mayor número de sesiones, que abarca las 15 categorías. Este proceso conlleva un alto costo computacional debido al gran número de revisiones. La tabla 17 a continuación presenta los resultados para tres semanas, siendo la semana 3 la que contiene el mayor número de revisiones, con un total de 246,272 instancias. Para mitigar los costos computacionales, utilizaremos instancias solo del período entre el 02/01/2017 y el 08/01/2027. Además, la Semana 3 destaca con el mayor número de instancias en comparación con las otras semanas. Esto sugiere que puede haber una actividad de usuario significativa o interacciones de productos durante este período, lo que lo convierte en un marco de tiempo importante para analizar para nuestra tarea de recomendación basada en sesiones. Y este enfoque ofrece la ventaja de reducir el tiempo de procesamiento mientras se capturan datos relevantes.

Tabla 17. Análisis de instancias por semana.

Week 1		Week 2		Week3	
Date	**Instances**	**Date**	**Instances**	**Date**	**Instances**
26/12/2014	28,799	19/01/2016	35,690	02/01/2017	36,811
27/12/2014	27,031	20/01/2016	44,785	03/01/2017	42,204
28/12/2014	29,662	21/01/2016	35,769	04/01/2017	36,687
29/12/2014	40,000	22/01/2016	35,346	05/01/2017	37,025
30/12/2014	30,742	23/01/2016	29,830	06/01/2017	31,271
31/12/2014	29,930	24/01/2016	29,493	07/01/2017	31,385
01/12/2014	33,587	25/01/2016	34,529	08/01/2017	30,889
Total	219,751	-	245,442	-	246,272

En el segundo paso de preprocesamiento, se procesó los conjuntos de productos fusionándolos con el conjunto correspondiente de revisiones relacionadas con la misma categoría. Además, se extrae la información de categoría del campo *rank* usando expresiones regulares. Se eliminaron las sesiones con solo una interacción para garantizar la calidad del conjunto de datos. Cada conjunto preprocesado se guardó como un archivo JSON.

Finalmente, todos los conjuntos se combinaron para crear un conjunto de datos unificado que abarca todas las categorías y sesiones preprocesadas. Este conjunto de datos consolidado servirá como base para un análisis más detallado y el desarrollo de modelos en nuestra tarea de recomendación basada en sesiones.

6.4 Experimentos

En esta sección, se describe la utilización de Transformers4Rec [104] para idear estrategias de entrenamiento. Transformers4Rec sirve como un marco de trabajo de sistema de recomendación de extremo a extremo, que abarca las etapas de preprocesamiento de datos, entrenamiento del modelo y evaluación. Desarrollado en Python, el marco de trabajo aprovecha PyTorch y Transformers de Hugging Face para facilitar la implementación eficiente y la experimentación con modelos basados en transformers para tareas de recomendación.

La biblioteca NVIDIA NVTabular, estrechamente relacionada con Transformers4Rec, ofrece capacidades aceleradas por GPU para tareas de preprocesamiento. Esta herramienta admite técnicas de ingeniería de características diseñadas para recomendaciones basadas en sesiones, que incluyen operaciones para agrupar interacciones ordenadas por tiempo por usuario o sesión, así como truncar secuencias para conservar las primeras o últimas N interacciones. Además, NVTabular permite guardar los datos preprocesados en un formato Parquet estructurado y consultable. Una

de las principales ventajas de NVTabular es su capacidad para acelerar los procesos de entrenamiento y evaluación al cargar datos directamente en la GPU. Además, la biblioteca proporciona un archivo de configuración que permite a los usuarios especificar qué características deben tratarse como características continuas o categóricas, lo que mejora la flexibilidad y personalización para el entrenamiento del modelo.

Se implementa un enfoque de entrenamiento y evaluación incremental, en el cual se utiliza una ventana deslizante con una unidad de tiempo única, como un día u hora, en orden temporal para entrenar el modelo de forma incremental. Esto implica ajustar finamente los parámetros de un modelo que ya ha sido entrenado utilizando datos pasados. Las sesiones se dividen en ventanas de tiempo, denominadas como T, con cada ventana teniendo una longitud de un día. La evaluación se lleva a cabo para cada ventana de tiempo subsiguiente Ti + 1, donde i a n-1, utilizando sesiones de ventanas de tiempo pasadas para el entrenamiento [T1, ..., Ti]. Además, las sesiones dentro de cada ventana de tiempo se dividen en una proporción de 50:50 entre conjuntos de validación y prueba. Los conjuntos de validación de cada ventana de tiempo se utilizan para la sintonización de hiperparámetros, mientras que los conjuntos de prueba se emplean para informar las métricas. Finalmente, las métricas informadas son los promedios en todas las ventanas de tiempo.

Evaluation en recomendación basada en sesiones se realiza utilizando métricas de clasificación Top-N tradicionales como NDCG@N y Recall@N. Normalized Discounted Cumulative Gain (NDCG) mide la efectividad de un sistema de clasificación considerando la posición de los elementos relevantes en la lista clasificada. Toma en cuenta la noción de que los elementos más altos en la clasificación deberían recibir más crédito que los elementos más bajos en la clasificación. Recall@N, por otro lado, evalúa la proporción de elementos relevantes correctamente identificados en las principales N recomendaciones, en relación con el número total de elementos relevantes en el conjunto de datos. En términos más simples, indica cuántos elementos relevantes fueron identificados con éxito entre las principales N recomendaciones. Las fórmulas para estas métricas son las siguientes.

$$NDCG@N = \frac{DCG@N}{IDCG@N} = \frac{\sum_{i=1}^{k\ (actual\ order)} \frac{Gains}{log_2(i+1)}}{\sum_{i=1}^{k\ (ideal\ order)} \frac{Gains}{log_2(i+1)}}$$

$$Recall@N = \frac{No.of\ recommended\ items\ @N\ that\ are\ relevant}{Total\ No.of\ relevant\ items}$$

En nuestros experimentos, utilizamos la arquitectura del transformador XLNet, que ofrece soporte tanto para modelado de lenguaje auto-regresivo como para auto-codificación, junto con la estrategia de entrenamiento PLM. Además, XLNet también puede emplear otras estrategias de entrenamiento como MLM, CLM y RTD. Realizamos dos experimentos utilizando diferentes conjuntos de datos. El primer conjunto de datos se centró en un dominio específico de productos de Amazon conocido como Amazon Fashion. El segundo conjunto de datos abarcó 15 dominios diversos como se describe en la Tabla 13, que se detalló en la sección anterior.

En el primer experimento, se enfoca en un solo dominio de Amazon Fashion, que contenía 186,637 instancias. Comenzamos preprocesando el conjunto de datos, lo que implicó eliminar columnas no utilizadas tanto de las reseñas como de los metadatos del producto. Además, se eliminaron las instancias duplicadas y se agregó una nueva columna llamada *event_type* con un valor de *purchase* para indicar el tipo de interacción. Los conjuntos de productos y reseñas se fusionaron. Se asignaron IDs de sesión concatenando el *reviewerID* y *unixReviewTime*, y se eliminaron las sesiones con una sola interacción. Este enfoque de preprocesamiento fue similar al utilizado para el conjunto de datos multidominio. Se definieron características categóricas como *user_id, event_type, brand, category, verified, price* y *overall*. Se extrajeron características temporales basadas en *unixReviewTime* para crear características cíclicas (seno y coseno) que pudieran representarse en un espacio continuo. Además, se normalizaron características continuas como *overall*. Sin embargo, se descubrió que *overall* era más útil como característica categórica. La longitud máxima de la sesión se estableció en 20, y el modelo se entrenó utilizando un enfoque de ventana deslizante. Específicamente, el modelo se entrenó con datos de cuatro días de cada siete días, con datos de validación del día siguiente. Este proceso continuó de manera iterativa, con el conjunto de entrenamiento avanzando un día a la vez. Se aplicaron cuatro estrategias de entrenamiento diferentes: MLM, PLM, RTD y CLM. Además, se realizó un sub experimento para explorar el uso de información adicional para la predicción del siguiente ítem, incorporando características categóricas.

En el segundo experimento, se siguió un enfoque similar al primer experimento, pero esta vez utilizamos un conjunto de datos multidominio que contiene productos de 15 categorías diferentes en Amazon. Los pasos de preprocesamiento fueron similares a los del primer experimento. Se definieron características categóricas como *user_id, event_type, brand, category, verified, price* y *overall*, y se extrajeron características temporales de *unixReviewTime*. Las características continuas se normalizaron, con *overall* siendo tratado principalmente como una característica categórica. La longitud máxima de la sesión se estableció en 20, y el modelo se entrenó utilizando un enfoque de ventana deslizante similar al primer experimento. Sin embargo, en este experimento nos enfocamos únicamente en la estrategia de entrenamiento MLM. Además, se realizó un sub experimento para explorar la integración de información adicional para la

predicción del siguiente ítem, incorporando características categóricas. Un desafío notable encontrado en este experimento fue el costo computacional del preprocesamiento de datos, debido al gran número de reseñas y productos asociados con las 15 categorías de productos de Amazon. Este desafío requirió una gestión cuidadosa de los recursos computacionales y la optimización de los flujos de trabajo de preprocesamiento para garantizar una ejecución eficiente.

Finalmente, se empleó una GRU entrenado con la estrategia CLM como modelo base. La elección de GRU, un tipo de Red Neuronal Recurrente (RNN), se realizó para facilitar una comparación entre las estrategias de entrenamiento basadas en transformers y aquellas basadas en RNN. Esta comparación tiene como objetivo evaluar la efectividad de las estrategias basadas en transformers en tareas de recomendación basadas en sesiones en relación con los enfoques tradicionales basados en RNN.

6.5 Resultados

Los resultados del primer experimento se representan en la Figura 25. Es evidente que las estrategias de entrenamiento de MLM, CLM, PLM y RTD basadas en la arquitectura XLNet superan al algoritmo GRU. Además, al incorporar información adicional para la predicción del próximo elemento utilizando características categóricas, la estrategia CLM logra los mejores resultados. Específicamente, esta estrategia logra una mejora notable del +168.81% tanto en NDCG@10 como en NDCG@20 en comparación con el valor base. Además, logra una mejora notable del +25% tanto en Recall@10 como en Recall@20 en comparación con el valor base. Estos resultados destacan la efectividad de las estrategias basadas en transformadores, especialmente CLM con información adicional, en mejorar el rendimiento de los sistemas de recomendación basados en sesiones.

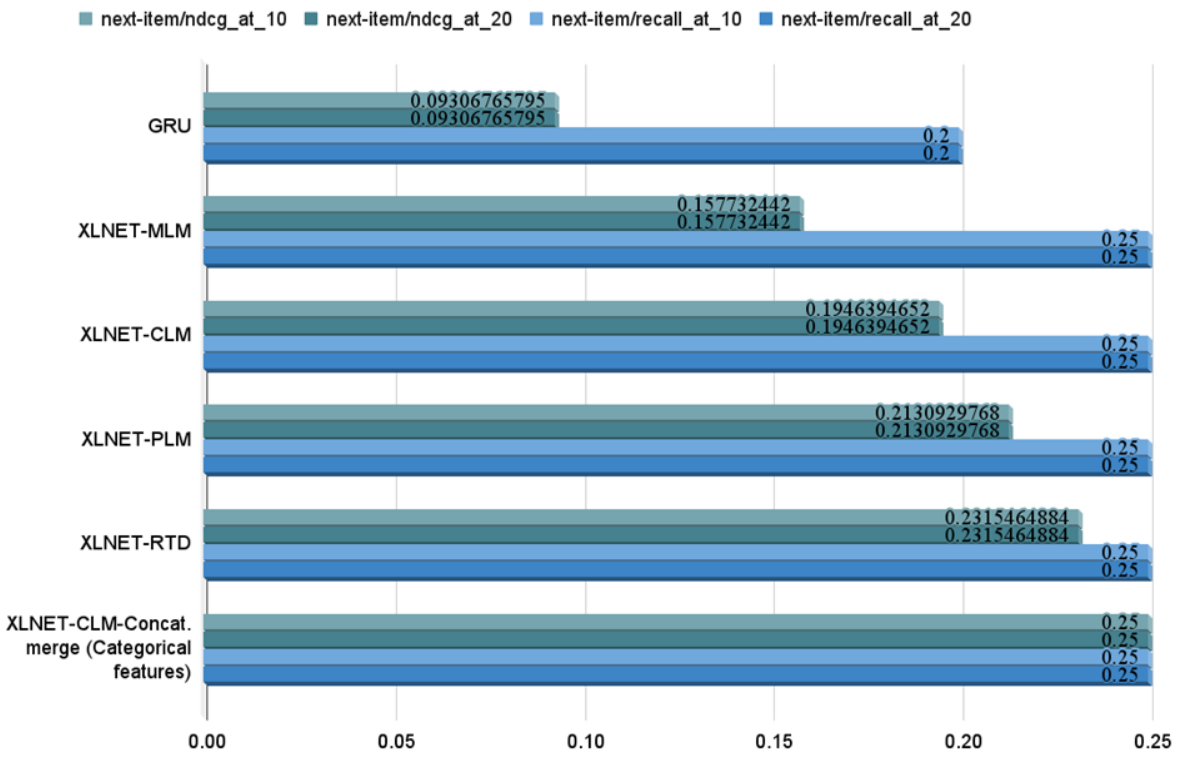

Figura 25. Los resultados del primer experimento con datos de un dominio específico.

Los resultados del segundo experimento se presentan en la Figura 26. Es evidente que la estrategia de entrenamiento de MLM basada en la arquitectura XLNet supera al algoritmo GRU. Además, al incorporar información adicional para la predicción del siguiente elemento utilizando características tanto categóricas como continuas, la estrategia de MLM logra los mejores resultados. Específicamente, esta estrategia logra una mejora significativa del +135.71% en NDCG@10 y del +136.23% en NDCG@20 en comparación con el modelo base. Además, alcanza una notable mejora del +86.39% en Recall@10 y del +95.69% en Recall@20 en comparación con el modelo base. Estos resultados destacan la superioridad de las estrategias basadas en transformadores, especialmente MLM con información adicional, en la mejora del rendimiento de los sistemas de recomendación basados en sesiones.

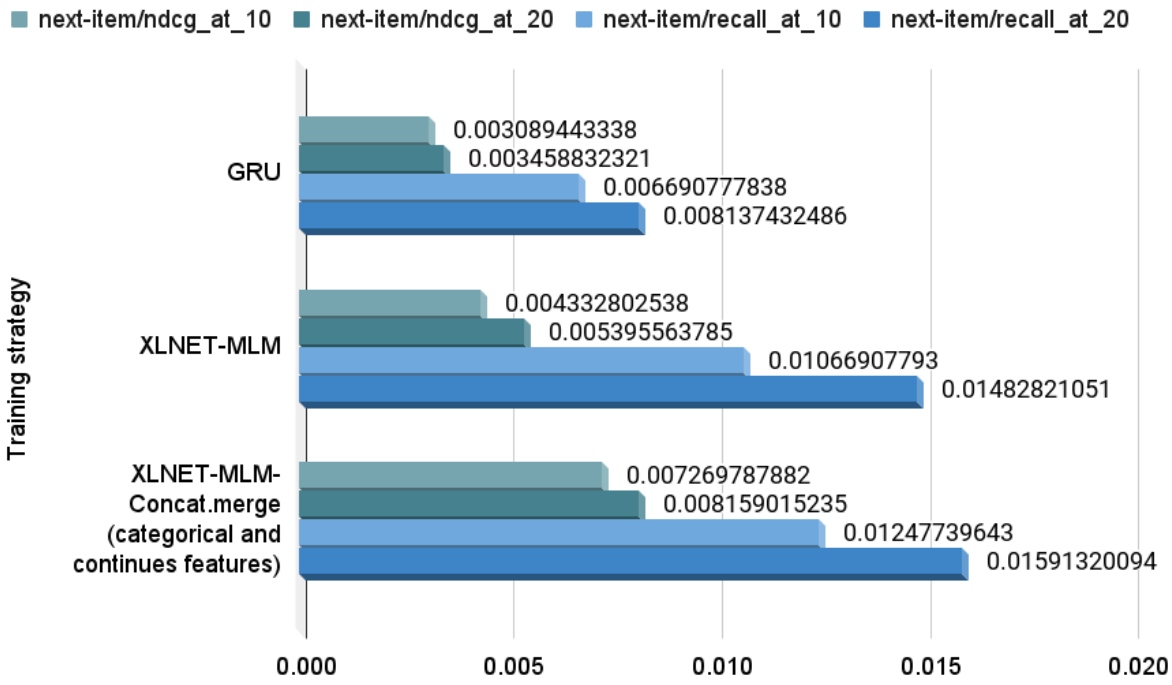

Figura 26. Los resultados del segundo experimento con datos de múltiples dominios.

6.6 Conclusiones

Se realizó un análisis sobre el un sistema de recomendación multidominio basado en Transformer para el e-commerce utilizando un conjunto de datos que comprende 102 millones de reseñas y metadatos de 6 millones de productos. Este conjunto de datos se sometió a un preprocesamiento para generar dos nuevos conjuntos de datos, uno para datos específicos de un dominio y otro para datos de múltiples dominios. En cada conjunto de datos, seleccionamos una semana con el mayor número de instancias para la experimentación.

Utilizando NVTabular, generamos características y seleccionamos características categóricas, continuas y cíclicas, como la calificación general, el estado de verificación, la categoría y las características temporales. Se aplicaron diversas estrategias de entrenamiento, incluido MLM, CLM, PLM y RTD, a la arquitectura de transformers XLNet. Estos modelos se entrenaron y evaluaron incrementalmente utilizando ventanas deslizantes por día.

Los resultados de estos experimentos se compararon con los obtenidos de un modelo de RNN, específicamente GRU. Los hallazgos demostraron que la arquitectura de transformers XLNet, junto con diversas estrategias de entrenamiento, superó a GRU tanto en datos de un solo dominio como en datos de múltiples dominios. Esto resalta la eficacia de los enfoques basados en transformers en sistemas de recomendación

basados en sesiones para datos multidominio. Sin embargo, es importante señalar que, si bien se observaron mejoras notables en los experimentos con datos multidominio, los resultados fueron ligeramente mejores para los datos de un dominio específico. Esto sugiere que la tarea de los sistemas de recomendación basados en sesiones para datos multidominio sigue siendo desafiante y requiere una investigación adicional.

Agradecimientos

Los autores expresan su gratitud al Laboratorio de Ingeniería de Lenguaje y Conocimiento (LKE) y la Facultad de Ciencias de la Computación (FCC) de la Benemérita Universidad Autónoma de Puebla (BUAP), School of Enterprise Computing and Digital Transformation of Technological University Dublin (TU Dublin), y el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por su apoyo al proporcionar recursos informáticos y asistencia financiera.

Capítulo 7: Conclusiones y trabajos futuros

En la era digital los sistemas de recomendación son una de las principales aplicaciones de ciencia de datos e inteligencia artificial para el ser humano debido a que se han integrado en nuestra vida diaria y son poderosas herramientas para la toma de decisiones informadas, efectivas y eficientes. Esta tecnología aborda la sobrecarga de información que es un problema que crece cada día en un mundo globalizado en donde más del 60% de la población es un usuario de internet que navega en promedio 6 horas diarias en el ciberespacio. De esta manera las compañías, instituciones u otras organizaciones se ven obligados a usar dentro de sus principales servicios los sistemas de recomendación para mejorar la experiencia de sus usuarios y reducir la tasa de abandono de sus usuarios en sus plataformas web como es el caso del comercio electrónico. El e-commerce es una actividad que es fuertemente impulsada por la naturaleza de la era digital y recientemente por la pandemia causada por el virus SARS-CoV-2. La pandemia del COVID-19 provocó que múltiples países aceleraran sus procesos de digitalización y cambiaran sus canales de venta al comercio electrónico como fue el caso en México reportado por la Asociación Mexicana de Venta Online (AMVO).

En este trabajo se realizó una investigación aplicada y experimental con un enfoque cuantitativo sobre sistemas de recomendación híbridos para el comercio electrónico. Se desarrolló una revisión sistemática de la literatura sobre enfoques híbridos de recomendación para identificar sus principales tendencias de desafíos, metodologías (algoritmos, técnicas, modelos y experimentos), dominios de aplicación, conjuntos de datos, formas de evaluación y áreas de oportunidad de investigación. Así mismo, se generó un modelo híbrido de recomendación basado en recuperación de información para textos turísticos mexicanos en el foro del Rest-Mex 2022, en el cual un modelo espacio vectorial fue desarrollado usando un esquema de ponderación tf-idf y la similitud del coseno. El filtrado colaborativo basado en el ítem y el filtrado basado en contenido se hibridaron con el enfoque por selección para abordar el problema del

arranque en frío e incrementar la efectividad de las recomendaciones. Este modelo híbrido de recomendación logro el segundo y tercer mejor resultado en el ranking de la tarea de sistema de recomendación del Rest-Mex 2022 y presentó una mínima diferencia de 0.0048 de MAE con respecto al resultado del equipo ganador.

Adicionalmente, se realizó una investigación sobre el análisis de sentimientos debido a que este tipo de tecnología puede retroalimentar un sistema de recomendación para mejorar la calidad de sus predicciones, es decir, al analizar el sentimiento de una reseña de un usuario se puede utilizar la polaridad del sentimiento positivo o negativo para generar recomendaciones más precisas. En este estudio se abordó la tarea de análisis de sentimiento del Rest-Mex 2023, la cual consistía en la identificación de la polaridad, el tipo y el país de una opinión dada. Estas subtareas se resolvieron con diversas estrategias de entrenamiento para el transformer RoBERTa como un ajuste fino, una adaptación de dominio y una técnica de aumento de datos. Nuestro equipo llamado LKE-IIMAS logró el primer lugar en el ranking de la tarea de análisis de sentimiento del Rest-Mex 2023.

Finalmente, se analizaron los sistemas de recomendación basados en sesiones para el e-commerce utilizando los últimos avances de Procesamiento de Lenguaje Natural (NLP) como son las arquitecturas transformer debido a que las técnicas tradicionales de recomendación como el filtrado colaborativo, el filtrado basado en contenido y los enfoques híbridos no consideran las preferencias dinámicas y a corto plazo del usuario. Para dirigir estas limitaciones se genero un sistema de recomendación basado en sesiones usando XLNet transformer con diversas estrategias de entrenamiento basada en el modelado del lenguaje como es MLM, CLM, PLM y RTD. Un conjunto de datos con 102 millones de reseñas de productos de Amazon fue pre-procesado para crear dos nuevos conjuntos, un conjunto para un solo dominio de productos y otro conjunto para múltiples dominios de productos. Se realizo una comparación con una red neuronal recurrente GRU, en donde las estrategias de entrenamiento con el transformer XLNet lograron una mejor eficiencia en sus recomendaciones que la red neuronal recurrente GRU.

Los trabajos futuros son en una primera instancia hibridar y evaluar otras técnicas de recomendación basadas en el procesamiento de lenguaje natural. Así mismo, analizar los sistemas de recomendación basados en sesiones con otras arquitecturas de Transformers y otros algoritmos de aprendizaje profundo. Una tendencia novedosa en este campo son los sistemas de recomendación conversacionales, los cuáles son una gran área de oportunidad para realizar una estancia posdoctoral sobre un dominio de comercio electrónico o educación.

7.1 Producción académica

A continuación, se presenta una lista de los productos académicos logrados con base en este trabajo de tesis.

7.1.1 Artículos de revista

- **Morales, V.G**., Pinto, D.E., Rojas, F., Gonzales, J.M.: A Systematic Literature Review on the Hybrid Approaches for Recommender Systems. Computacion y Sistemas. 26, 357–372 (2022). https://doi.org/10.13053/CyS-26-1-4180.
- **Morales-Murillo, V.G**., Pinto, D., Perez-Tellez, F., Rojas-Lopez, F.: A Transformer-Based Multi-Domain Recommender System for E-commerce. International Journal of Combinatorial Optimization Problems and Informatics. 15, 95–123 (2024). https://doi.org/10.61467/2007.1558.2024.v15i2.465.

7.1.2 Artículos de conferencia

- **Morales, V.G**., Pinto, D.E., Rojas, F.: A Hybrid Recommender Model based on Information Retrieval for Mexican Tourism Text in Rest-Mex 2022. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022). CEUR Workshop Proceedings, A Coruña, Spain (2022).
- **Morales-Murillo, V.G.**, Gómez-Adorno, H., Pinto, D., Cortés-Miranda, I.A., Delice, P.: LKE-IIMAS team at Rest-Mex 2023: Sentiment Analysis on Mexican Tourism Reviews using Transformer-Based Domain Adaptation. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), Jaén, Spain (2023).

7.1.3 Capítulos de libro

- **Morales, V.G.**, Pinto, D.E., Rojas, F.: Un modelo de recomendación basado en recuperación de información para textos turísticos mexicanos en español. In: Avances de ingeniería del lenguaje, del conocimiento y la interacción humano máquina Volumen II. pp. 59–67. United Academic Journals (UA Journals) (2022).

7.1.4 Estancias de investigación

- En el periodo primavera 2023 se realizó una estancia de investigación nacional en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM) bajo la supervisión de la Dra. Helena Montserrat Gómez Adorno.
- En el periodo Otoño 2023 se realizó una estancia de investigación internacional en la School of Enterprise Computing and Digital Transformation de la Technological University Dublin (TU Dublin) – Tallaght Campus bajo la supervisión del Dr. Fernando Pérez Tellez.

7.1.5 Premios

- 1° Lugar en la tarea Sentiment analysis for mexican touristic places del Foro Internacional Rest-Mex 2023: Research on Sentiment Analysis Task for Mexican Tourist Texts del IberLEF 2023.
- 2° Lugar en la tarea Recommendation system for mexican touristic places del Foro Internacional Rest-Mex 2022: Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts del IberLEF 2022.
- 1° Lugar en la sesión de presentaciones del Taller de Desarrollo de Ideas de Negocio Basadas en Inteligencia Artificial (INEBIA) colocado en el Congreso de Soluciones Industriales Basadas en Inteligencia Artificial (SIBIA 2021).

7.1.6 Talleres

- Impartición del Reto de Análisis de Sentimientos de Reseñas de Sitios Turísticos del 26 al 30 de junio, como parte del taller Macroentrenamiento en Inteligencia Artificial 2023 organizado por la Universidad Nacional Autónoma de México (UNAM) y la Red de Macrouniversidades de América Latina y el Caribe.
- Impartición del Tutorial Machine Learning para Clasificación de Textos del 19 al 23 de junio, como parte del taller Macroentrenamiento en Inteligencia Artificial 2023 organizado por la Universidad Nacional Autónoma de México (UNAM) y la Red de Macrouniversidades de América Latina y el Caribe.
- Impartición del Taller/MasterClass de Sistema de Recomendación Basado en Lenguaje Natural en el Tercer Foro Nacional de Tecnologías de la Información y Sistemas Computacionales durante el 28 y 29 de septiembre del 2023.

- Impartición del taller de Introducción a los Sistemas de Recomendación en el 5to Encuentro de Enfoques Tecnológicos Aplicados a Sistemas Computacionales en la Universidad Politécnica Metropolitana de Puebla (UPMP) durante el 8 y 9 de septiembre del 2022.

7.1.7 Presentaciones

- Presentación de artículo en el 9th International Symposium on Language & Knowledge Engineering (LKE 2024) que se realizó del 4 al 6 de junio en la Universidad Tecnológica de Dublín (TU Dublin), en Dublín, Irlanda.
- Presentación de 2 posters en el Primer Encuentro Nacional de Ciencia de Datos realizado del 9 al 10 de noviembre del 2023 en Puebla, México.
- Presentación de artículo en el Workshop del Iberian Languages Evaluation Forum (IberLEF 2023) que fue colocado dentro del marco del XXXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2023) que se realizó del 26 al 29 de septiembre en Jaén, España.
- Presentación de artículo en el Workshop del Iberian Languages Evaluation Forum (IberLEF 2022) que fue colocado dentro del marco del XXXVIII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2022) que se realizó del 19 al 23 de septiembre en La Coruña, España.
- Presentación de artículo en el foro del MexLEF 2022 que se realizó durante los días 6 y 7 de octubre en San Pedro Cholula, Puebla.

7.1.7 Colaboración extra

- 2° Lugar en la subtarea 2 (Subtask 2: Model Attribution) del AUtomated TEXt IdenTIFICATION on languages of the Iberian Peninsula (IberAuTexTification 2024) del IberLEF 2024.
- 4° Lugar en la subtarea 1 (Subtask 1: Human or Generated) del AUtomated TEXt IdenTIFICATION on languages of the Iberian Peninsula (IberAuTexTification 2024) del IberLEF 2024.

Referencias

1. Vaidya, N., A.R, K.: Recommender Systems-The need of the Ecommerce Era. In: 2017 International Conference on Computing Methodologies and Communication (ICCMC). pp. 100–104. IEEE (2017).

2. Jannach, D., Resnick, P., Tuzhilin, A., Zanker, M.: Recommender Systems — Beyond Matrix Completion. Commun ACM. 59, 94–102 (2016).

3. Nelson, B.S.: Spotify ' s Collaborative Filtering. 1–7.

4. INEGI: EL INEGI PRESENTA RESULTADOS DE LA SEGUNDA EDICIÓN DEL ECOVID-IE Y DEL ESTUDIO SOBRE LA DEMOGRAFÍA DE LOS NEGOCIOS 2020. (2020).

5. Asociación mexicana de venta online (AMVO): Estudio sobre Venta Online en PyMEs 2021. (2021).

6. Morales, V., Pinto, D., Vilariño, D.: Un nuevo paradigma basado en recuperación de información para sistemas de recomendación, (2020).

7. Mohamed, M.H., Khafagy, M.H., Ibrahim, M.H.: Recommender Systems Challenges and Solutions Survey. Proceedings of 2019 International Conference on Innovative Trends in Computer Engineering, ITCE 2019. 149–155 (2019). https://doi.org/10.1109/ITCE.2019.8646645.

8. Kemp, S.: Digital 2020. , New York, NY, USA (2020).

9. Tovar Garrido, L.C., Jimenez Garcia, E.E., Montoya Campo, J.D.: Sistema Ecléctico De Filtrado De Información Basado En Inteligencia Computacional Para Recomendación De Atractivos Turísticos Del Caribe Colombiano. Loginn. 1, 101 (2017).

10. Hernández, R., Fernández, C., Baptista, P.: Metodología de la investigación. McGrawHill, México, D.F. (2014).

11. Ñaupas, H., Paitán, Marcelino Raúl Valdivia Dueñas, Jesús Josefa Palacios Vilela, H.E.R.D.: Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis. (2018). https://doi.org/10.1017/CBO9781107415324.004.

12. Londoño, O., Maldonado, L., Calderón, L.: Guía para construir estados del arte. Internacional Corporation of Network of Knowledge (ICONK), Bogotá (2016).

13. Herlocker, J., Konstan, J., Terveen, L., Riedl, J.: Evaluating Collaborative Filtering Recommender Systems. ACM Trans Inf Syst. 22, 5–53 (2004).

14. Borges, H.L., Lorena, A.C.: A survey on recommender systems for news data. (2010). https://doi.org/10.1007/978-3-642-04584-4_6.

15. Costa, A., Roda, F.: Recommender systems by means of information retrieval. ACM International Conference Proceeding Series. (2011). https://doi.org/10.1145/1988688.1988755.

16. Stanescu, A., Nagar, S., Caragea, D.: A Hybrid Recommender System: User Profiling from Keywords and Ratings. In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). pp. 73–80. IEEE (2013).

17. Isinkaye, F., Folajimi, Y., Ojokoh, B.: Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal. 16, 261–273 (2015).

18. Gabrani, G., Sabharwal, S.: Artificial Intelligence Based Recommender Systems: A Survey. In: ICACDS 2016. pp. 50–29. Springer, Ghaziabad (2016).

19. Escamilla González, O., Marcellin Jacques, S.: Estado del arte en los sistemas de recomendación. Research in Computing Science. 135, 25–40 (2017). https://doi.org/10.13053/rcs-135-1-2.

20. Gómez, P., Guarda, T., Cedeño, J., Benavides, A., Alejandro, C., Mosquera, G., Garcia, T., Benavides, V.: Sistemas de Recomendación: un enfoque a las técnicas de filtrado. Revista Ibérica de Sistemas e Tecnologias de Informação. 286–293 (2019).

21. Rahul, Dahiya, H., Singh, D.: A Review of Trends and Techniques in Recommender Systems. In: 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). pp. 1–8. IEEE, Ghaziabad (2019).

22. Artemenko, O., Pasichnyk, V., Kunanec, N.: E-tourism mobile location-based hybrid recommender system with context evaluation. In: 14th International Conference on Computer Sciences and Information Technologies (CSIT). pp. 114–118. IEEE, Lviv (2019).

23. Zhao, X.: A Study on E-commerce Recommender System Based on Big Data. In: 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). pp. 222–226. IEEE, Chengdu (2019).

24. Idrissi, N., Zellou, A., Hourrane, O., Bakkoury, Z., Benlahmar, E.H.: A New Hybrid-Enhanced Recommender System for Mitigating Cold Start Issues. In: ICIME 2019: Proceedings of the 2019 11th International Conference on Information Management and Engineering. pp. 10–14. Association for Computing Machinery, New York, NY, USA (2019).

25. Najmani, K., El habib, B., Sael, N., Zellou, A.: A Comparative Study on Recommender Systems Approaches. In: Proceedings of the 4th International Conference on Big Data and Internet of Things. pp. 1–5. Association for Computing Machinery, New York, NY, USA (2019).

26. Duan, L., Gui, X., Wei, M., Wei, M.: A Resume Recommendation Algorithm Based on K-means++ and Part-of-speech TF-IDF. In: Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing. pp. 1–5. Association for Computing Machinery, New York, NY, USA (2019).

27. Goyani, M., Chaurasiya, N.: A Review of Movie Recommendation System: Limitations, Survey and Challenges. Electronic Letters on Computer Vision and Image Analysis. 19, 18–37 (2020). https://doi.org/10.5565/rev/elcvia.1232.

28. Banik, R.: Hands-On Recommendation Systems with Python. Packt, Birmingham, mumbai (2018).

29. Chen, M., Liu, P.: Performance Evaluation of Recommender Systems. International Journal of Performability Engineering. 13, 1246–1256 (2017).

30. Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts: Working Papers on Information Systems. 10, 10–26 (2012).

31. Çano, E., Morisio, M.: Hybrid Recommender Systems: A Systematic Literature Review. Intelligent Data Analysis. 21, 1487–1524 (2017).

32. Malekpour Alamdari, P., Jafari Navimipour, N., Hosseinzadeh, M., Ali Safaei, A., Darwesh, A.: A Systematic Study on the Recommender Systems in the E-Commerce. IEEE Access. 8, 115694–115716 (2020).

33. Mohammadi, V., Masoud Rahmani, A., Mohammed Darwesh, A., Sahaf, A.: Trust-based recommendation systems in Internet of Things: a systematic literature review. Human-centric Computing and Information Sciences. 9, 1–61 (2019).

34. Katarya, R., Arora, Y.: Capsmf: a novel product recommender system using deep learning based text analysis model. Multimed Tools Appl. 79, 35927–35948 (2020).

35. Jabbar, M., Javaid, Q., Arif, M., Munir, A., Javed, A.: An Efficient and Intelligent Recommender System for Mobile Platform. MEHRAN UNIVERSITY RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY. 37, 463–480 (2018).

36. Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., Getoor, L.: Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. ACM Trans Interact Intell Syst. 10, (2020).

37. Ifada, N., Ummamah, Kautsar, M.: Hybrid Popularity Model for Solving Cold-start Problem in Recommendation System. In: SIET '20: Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology. pp. 40–44. Association for Computing Machinery, Malang, Indonesia (2020).

38. Do, H.Q., Le, T.H., Yoon, B.: Dynamic Weighted Hybrid Recommender Systems. In: 2020 22nd International Conference on Advanced Communication Technology (ICACT). pp. 644–650 (2020).

39. Idrissi, N., Zellou, A., Hourrane, O., Bakkoury, Z., Benlahmar, E.H.: Addressing Cold Start Challenges in Recommender Systems: Towards A New Hybrid Approach. In: 2019 International Conference on Smart Applications, Communications and Networking (SmartNets). pp. 1–6. IEEE, Sharm El Sheikh (2019).

40. Tahmasebi, F., Meghdadi, M., Ahmadian, S., Valiallahi, K.: A hybrid recommendation system based on profile expansion technique to alleviate cold start problem. Multimed Tools Appl. 80, 2339–2354 (2021).

41. Devika, R., Subramaniyaswamy, V.: A novel model for hospital recommender system using hybrid filtering and big data techniques. In: 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on. pp. 575–579. IEEE (2018).

42. Cao, D., He, X., Nie, L., Wei, X., Hu, X., Wu, S., Chua, T.S.: Cross-Platform App Recommendation by Jointly Modeling Ratings and Texts. ACM Trans Inf Syst. 35, (2017).

43. Oh, J., Kim, S., Yun, S.Y., Choi, S., Yong Yi, M.: A Pipelined Hybrid Recommender System for Ranking the Items on the Display. In: RecSys Challenge '19: Proceedings of the Workshop on ACM Recommender Systems Challenge. pp. 1–5. Association for Computing Machinery, New York, NY, USA (2019).

44. Kavitha, S., Rajesh Badre, R.: Towards a Hybrid Recommendation System on Apache Spark. In: 2020 IEEE India Council International Subsections Conference (INDISCON). pp. 297–302. IEEE (2020).

45. Troussas, C., Krouska, A., Virvou, M.: Multi-algorithmic techniques and a hybrid model for increasing the efficiency of recommender systems. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). pp. 184–188. IEEE (2018).

46. Yang, Y., Jang, H., Kim, B.: A Hybrid Recommender System for Sequential Recommendation : Combining Similarity Models With Markov Chains. IEEE Access. 8, 190136–190146 (2020).

47. Hernández-Rubio, M., Cantador, I., Bellogín, A.: A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. User Model User-adapt Interact. 29, 381–441 (2019).

48. WAIREGI, S., MWANGI, W., RIMIRU, R.: A Framework for Items Recommendation System Using Hybrid Approach. In: 2020 IST-AFRICA CONFERENCE (IST-AFRICA). pp. 1–15 (2020).

49. A. Heggo, I., Abdelbaki, N.: Hybrid Information Filtering Engine for Personalized Job Recommender System. In: Ella Hassanien, A., F. Tolba, M., Elhoseny, M., and Mostafa, M. (eds.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018). pp. 553–563. Springer, Cham (2018).

50. Zheng, X., Luo, Y., Sun, L., Ding, X., Zhang, J.: A novel social network hybrid recommender system based on hypergraph topologic structure. WORLD WIDE WEB-INTERNET AND WEB INFORMATION SYSTEMS. 21, 985–1013 (2018).

51. Li, X., Chen, Y., Pettit, B., De Rijke, M.: Personalised Reranking of Paper Recommendations Using Paper Content and User Behavior. ACM Trans Inf Syst. 37, 1–23 (2019).

52. Logesh, R., Subramaniyaswamy, V., Malathi, D., Sivaramakrishnan, N., Vijayakumar, V.: Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method. Neural Comput Appl. 32, 2141–2164 (2020).

53. Choenaksorn, S., Maneeroj, S.: New Location Recommendation Technique on Social Network. In: ICISS '18: Proceedings of the 2018 International Conference on Information Science and System. pp. 3–7. Association for Computing Machinery, New York, NY, USA (2018).

54. Naz, S., Maqsood, M., Yahya Durani, M.: An Efficient Algorithm for Recommender System Using Kernel Mapping Techniques. In: ICSCA '19: Proceedings of the 2019 8th International Conference on Software and

Computer Applications. pp. 115–119. Association for Computing Machinery, New York, NY, USA (2019).

55. Fatih Cetin, M., Ayvaz, S.: A Negative Similarity Based Hybrid Recommender System Using Apache Spark. In: ICAAI 2019: Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence. pp. 166–172. Association for Computing Machinery, New York, NY, USA (2019).

56. Mueller, J.: Combining Aspects of Genetic Algorithms with Weighted Recommender Hybridization. In: iiWAS '17: Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services. pp. 13–22. Association for Computing Machinery, Salzburg, Austria (2017).

57. Ferraro, A., Bogdanov, D., Yoon, J., Kim, K., Serra, X.: Automatic playlist continuation using a hybrid recommender system combining features from text and audio. In: RecSys Challenge '18: Proceedings of the ACM Recommender Systems Challenge 2018. pp. 1–5. Association for Computing Machinery, Vancouver, BC, Canada (2018).

58. Grad-gyenge, L., Kiss, A., Filzmoser, P.: Graph Embedding Based Recommendation Techniques on the Knowledge Graph. In: UMAP '17: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization. pp. 354–359. Association for Computing Machinery, Bratislava, Slovakia (2017).

59. Patel, A.A., N. Dharwa, J.: Fuzzy Based Hybrid Mobile Recommendation System. In: ICTCS '16: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. pp. 1–6. Association for Computing Machinery, Udaipur, India (2016).

60. Al-otaibi, S., Ykhlef, M.: Hybrid immunizing solution for job recommender system. Front Comput Sci. 11, 511–527 (2017).

61. Elyes Ben Haj Kbaier, M., Masri, H., Krichen, S.: A personalized hybrid tourism recommender system. In: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). pp. 244–250. IEEE (2017).

62. Duzen, Z., S. Aktas, M.: An Approach To Hybrid Personalized Recommender Systems. In: PROCEEDINGS OF THE 2016 INTERNATIONAL SYMPOSIUM ON INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS (INISTA). pp. 1–8. , Sinaia, ROMANIA (2016).

63. Waqar, M., Majeed, N., Dawood, H., Daud, A., Radi Aljohani, N.: An adaptive doctor-recommender system. Behaviour & Information Technology. 38, 959–973 (2019).

64. D. Nagowah, S., Rajarai, K.: A Hybrid Approach for Recommender Systems in a Proximity Based Social Network. In: 2nd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM). pp. 302–312. Springer International Publishing, MAURITIUS (2019).

65. Kaššák, O., Kompan, M., Bieliková, M.: Personalized Hybrid Recommendation for Group of Users : Top-N Multimedia Recommender. Inf Process Manag. 52, 459–477 (2016).

66. Ravi, L., Subramaniyaswamy, V., Vijayakumar, V., Chen, S., Karmel, A., Devarajan, M.: Hybrid Location-based Recommender System for Mobility and Travel Planning. MOBILE NETWORKS & APPLICATIONS. 24, 1226–1239 (2019).

67. Liu, Y., Yang, B., Pei, H., Huang, J.: Neural Explainable Recommender Model Based on Attributes and Reviews. J Comput Sci Technol. 35, 1446–1460 (2020).

68. Nilashi, M., Ibrahim, O., Bagherifard, K.: A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Syst Appl. 92, 507–520 (2018).

69. Wang, S., Hu, L., Wang, Y., Cao, L., Sheng, Q.Z., Orgun, M.: Sequential Recommender Systems : Challenges , Progress and Prospects ∗. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19). pp. 6332–6338. International Joint Conferences on Artificial Intelligence Organization, Macao (2019).

70. Kamath, U., Liu, J., Whitaker, J.: Deep Learning for NLP and Speech Recognition. Springer Nature Switzerland AG, Cham, Switzerland (2019).

71. Kochenderfer, M., Wheeler, T.: Algorithms for Optimization. The MIT Press, Cambridge, Massachusetts (2019).

72. Manning, C., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2009).

73. UNTWO: Glossary of tourism terms | UNTWO, https://www.unwto.org/glossary-tourism-terms, last accessed 2022/05/18.

74. UNWTO: Why Tourism?, https://www.unwto.org/why-tourism, last accessed 2022/05/18.

75. Álvarez, M.Á., Aranda, R., Arce, S., Fajardo, D., Guerrero, R., López, A.P., Martínez, J., Pérez, H., Rodríguez, A.Y.: Overview of Rest-Mex at IberLEF 2021 : Recommendation System for Text Mexican Tourism. Procesamiento del Lenguaje Natural. 67, (2021).

76. Álvarez, M.Á., Díaz, Á., Aranda, R., Rodríguez, A.Y., Fajardo, D., Guerrero, R., Bustio, L.: Overview of Rest-Mex at IberLEF 2022: Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts. Procesamiento del Lenguaje Natural. 69, (2022).

77. Morales, V.G., Pinto, D.E., Rojas, F., Gonzales, J.M.: A Systematic Literature Review on the Hybrid Approaches for Recommender Systems. Computacion y Sistemas. 26, 357–372 (2022). https://doi.org/10.13053/CyS-26-1-4180.

78. Idrissi, N., Zellou, A., Hourrane, O., Bakkoury, Z., Benlahmar, E.H.: A New Hybrid-Enhanced Recommender System for Mitigating Cold Start Issues. In: ICIME 2019: Proceedings of the 2019 11th International Conference on Information Management and Engineering. pp. 10–14. Association for Computing Machinery, New York, NY, USA (2019).

79. Arreola, J., Garcia, L., Ramos, J., Rodríguez, A.: An Embeddings Based Recommendation System for Mexican Tourism . Submission to the REST-MEX Shared Task at IberLEF 2021. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). pp. 110–117. , Málaga, España (2021).

80. Morales, E., Torres, D., Ehrlich, A., Toledo, M., Martínez, B., Hermosillo, J.: A Recommendation System for Tourism Based on Semantic Representations and Statistical Relational Learning. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). pp. 134–148. , Málaga, España (2021).

81. flaticon.com: Iconos vectoriales y stickers - PNG, SVG, EPS, PSD y CSS, https://www.flaticon.es/, last accessed 2023/01/23.

82. Géron, A.: Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly, CA 95472 (2019).

83. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems An Introduction. Cambridge University Press, New York, NY, USA (2011).

84. Álvarez-Carmona, M.Á., Díaz-Pacheco, Á., Ramón, A., Rodríguez-González, A.Y., Bustio-Martínez, L., Muñis-Sánchez, V., Pastor-López, A.P., Sánchez-Vega, F.: Overview of Rest-Mex at IberLEF 2023: Research on Sentiment Analysis Task for Mexican Tourist Texts. Procesamiento del Lenguaje Natural. 71, (2023).

85. CIMAT: Rest-Mex: Research on Sentiment Analysis Task for Mexican Tourist Texts, https://sites.google.com/cimat.mx/rest-mex2023/, last accessed 2023/05/21.

86. Balcí, S., Demirci, G.M., Demirhan, H., Sarp, S.: Sentiment Analysis Using State of the Art Machine Learning Techniques. In: Biele, C., Kacprzyk, J., Kopeć, W., Owsiński, J.W., Romanowski, A., and Sikorski, M. (eds.) Proceedings of MIDI'2021 – 9th Machine Intelligence and Digital Interaction Conference. pp. 34–42. Springer International Publishing, Warsaw (2022).

87. Vásquez, J., Gómez-Adorno, H., Bel-Enguix, G.: Bert-based Approach for Sentiment Analysis of Spanish Reviews from TripAdvisor. In: Proceedings of the Third Workshop for Iberian Languages Evaluation Forum (IberLEF 2021). pp. 163–172. CEUR WS Proceedings (2021).

88. Ballester-Espinosa, A.: Cascade of Biased Two-class Classifiers for Multi-class Sentiment Analysis. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2021). CEUR-WS.org, Málaga (2021).

89. Velazquez Medina, G., Hernández Farías, D.I.: DCI-UG participation at REST-MEX 2021: A Transfer Learning Approach for Sentiment Analysis in Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2021). CEUR-WS.org, Málaga (2021).

90. y Ángel Díaz-Pacheco y Ramón Aranda y Ansel Y. Rodríguez-González y Daniel Fajardo-Delgado y Rafael Guerrero-Rodríguez y Lázaro Bustio-Martínez, M.Á.Á.-C.: Overview of Rest-Mex at IberLEF 2022: Recommendation System, Sentiment Analysis and Covid Semaphore Prediction for Mexican Tourist Texts. Procesamiento del Lenguaje Natural. 69, 289–299 (2022).

91. García-Díaz, J.A., Rodríguez-García, M.Á., García-Sánchez, F., Valencia-García, R.: UMUTeam at REST-MEX 2022: Polarity Prediction using Knowledge Integration of Linguistic Features and Sentence Embeddings based on Transformers. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022). CEUR Workshop Proceedings (CEUR-WS.org), A Coruña, Spain (2022).

92. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, \Lukasz, Polosukhin, I.: Attention is All You Need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6000–6010. Curran Associates Inc., Red Hook, NY, USA (2017).

93. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv. abs/1810.04805, (2019).

94. Katharopoulos, A., Vyas, A., Pappas, N., Fleuret, F.: Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention. In: International Conference on Machine Learning (2020).

95. Tunstall, L., von Werra, L., Wolf, T.: Natural Language Processing with Transformers, Revised Edition. O'Reilly Media (2022).

96. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach. CoRR. abs/1907.11692, (2019).

97. Mohammed, R., Rawashdeh, J., Abdullah, M.: Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. In: 2020 11th International Conference on Information and Communication Systems (ICICS). pp. 243–248 (2020). https://doi.org/10.1109/ICICS49469.2020.239556.

98. Pham, N.L., Vinh Nguyen, V., Pham, T.V.: A Data Augmentation Method for English-Vietnamese Neural Machine Translation. IEEE Access. 11, 28034–28044 (2023). https://doi.org/10.1109/ACCESS.2023.3252898.

99. Zhu, Y., Qiu, Y., Wu, Q., Wang, F.L., Rao, Y.: Topic Driven Adaptive Network for cross-domain sentiment classification. Inf Process Manag. 60, (2023). https://doi.org/10.1016/j.ipm.2022.103230.

100. Gutiérrez-Fandiño, A., Armengol-Estapé, J., Pàmies, M., Llop-Palao, J., Silveira-Ocampo, J., Carrino, C.P., Gonzalez-Agirre, A., Armentano-Oller, C., Rodriguez-Penagos, C., Villegas, M.: Spanish Language Models, (2021).

101. Wang, S., Pasi, G., Hu, L., Cao, L.: The Era of Intelligent Recommendation: Editorial on Intelligent Recommendation with Advanced AI and Learning. IEEE Intell Syst. 35, 3–6 (2020). https://doi.org/10.1109/MIS.2020.3026430.

102. Wang, S., Cao, L., Wang, Y., Sheng, Q.Z., Orgun, M.A., Lian, D.: A Survey on Session-based Recommender Systems. ACM Comput Surv. 54, (2022). https://doi.org/10.1145/3465401.

103. Jannach, D., Ludewig, M., Lerche, L.: Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. User Model User-adapt Interact. 27, 351–392 (2017). https://doi.org/10.1007/s11257-017-9194-1.

104. de Souza Pereira Moreira, G., Rabhi, S., Lee, J.M., Ak, R., Oldridge, E.: Transformers4Rec: Bridging the Gap between NLP and Sequential / Session-Based Recommendation. In: Proceedings of the 15th ACM Conference on Recommender Systems. pp. 143–153. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3460231.3474255.

105. Ni, J., Li, J., McAuley, J.: Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In: Inui, K., Jiang, J., Ng, V., and Wan, X. (eds.) Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 188–197. Association for Computational Linguistics, Hong Kong, China (2019). https://doi.org/10.18653/v1/D19-1018.

106. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q. V: XLNet: Generalized Autoregressive Pretraining for Language Understanding, (2020).

107. Meng, Y., Krishnan, J., Wang, S., Wang, Q., Mao, Y., Fang, H., Ghazvininejad, M., Han, J., Zettlemoyer, L.: Representation Deficiency in Masked Language Modeling, (2023).

108. Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., Huang, X.: Pre-trained models for natural language processing: A survey. Sci China Technol Sci. 63, 1872–1897 (2020). https://doi.org/10.1007/s11431-020-1647-3.

109. Wu, X., Varshney, L.R.: A Meta-Learning Perspective on Transformers for Causal Language Modeling, (2023).

110. Lu, K., Potash, P., Lin, X., Sun, Y., Qian, Z., Yuan, Z., Naumann, T., Cai, T., Lu, J.: Prompt Discriminative Language Models for Domain Adaptation. In: Naumann, T., Ben Abacha, A., Bethard, S., Roberts, K., and Rumshisky, A. (eds.) Proceedings of the 5th Clinical Natural Language Processing Workshop. pp. 247–258. Association for Computational Linguistics, Toronto, Canada (2023). https://doi.org/10.18653/v1/2023.clinicalnlp-1.30.

111. Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., Jiang, P.: BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer, (2019).

112. Gheewala, S., Xu, S., Yeom, S., Maqsood, S.: Exploiting deep transformer models in textual review based recommender systems. Expert Syst Appl. 235, 121120 (2024). https://doi.org/https://doi.org/10.1016/j.eswa.2023.121120.