Optimización de Recomendaciones de Rutas Turísticas Mediante Prefiltrado Contextual en Estructuras de Grafos.

Mariña Dolado Trigo[1 [0009−0007−9460−6298]], José L. Jorro-Aragoneses[1 [0000−0002−2917−3912]], Alejandro Bellogín[1 [0000−0001−6368−2510]], and Pablo Sánchez[2 [0000−0003−1792−1706]]

[1] Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid
marinna.dolado@estudiante.uam.es, jose.jorro@uam.es
alejandro.bellogin@uam.es
[2] Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid
psperez@icai.comillas.edu

Resumen La planificación de rutas turísticas en ciudades enfrenta el reto de la sobrecarga de información y de la alta dependencia del contexto dinámico, factores que los sistemas de recomendación tradicionales a menudo no captan eficazmente. En este trabajo, proponemos un sistema de recomendación de rutas basado en una estructura de grafos que integra información contextual diferenciada: dimensiones estáticas de los puntos de interés (POIs) y variables dinámicas de las interacciones de desplazamiento. Este modelo permite aplicar técnicas de prefiltrado contextual de forma eficiente, reduciendo el espacio de búsqueda y proporcionando una interpretabilidad inherente que supera a los enfoques opacos de *"caja negra"*. A través de una evaluación preliminar con el conjunto de datos Context Trails en Tokio, observamos que las restricciones contextuales influyen de manera determinante en la selección de POIs. Estos resultados, aunque iniciales, sientan las bases para desarrollar modelos de recomendación más realistas, trazables y explicables, y establecen una hoja de ruta para futuras mejoras en la personalización y la sostenibilidad de las recomendaciones turísticas.

Keywords: Sistemas de recomendación · Recomendación de rutas · Recomendación basada en contexto · Grafos · Predicción del siguiente punto de interés.

1. Introducción

La planificación de itinerarios turísticos en entornos urbanos enfrenta el desafío de la sobrecarga de información ante la vasta oferta de puntos de interés (*Points of Interest*, POIs). Si bien los sistemas de recomendación mitigan este problema mediante el filtrado de lugares, la generación de rutas difiere de otros dominios al estar intrínsecamente condicionada por el contexto dinámico del usuario, que incluye su ubicación, la meteorología o el tiempo disponible [6,12].

A pesar de la alta precisión que ofrecen los modelos basados en aprendizaje profundo, su naturaleza de *"caja negra"* plantea un obstáculo crítico para la interpretabilidad y la transparencia en la toma de decisiones [3,4,11]. Esta opacidad no solo limita la confianza del usuario final, sino que dificulta la justificación del por qué se ha sugerido un determinado destino turístico. Además, algunas de las soluciones propuestas en la literatura dificultan su extensibilidad al incorporar nuevas dimensiones contextuales que puedan enriquecer las recomendaciones.

Para abordar estas limitaciones, proponemos un sistema de recomendación de rutas turísticas basado en un grafo de interacciones enriquecido con contexto. Nuestra propuesta introduce un modelo de datos que desacopla la información contextual de los nodos (POIs) de la asociada a las aristas (desplazamientos), lo que permite aplicar restricciones mediante técnicas de prefiltrado de manera eficiente y reducir drásticamente el espacio de búsqueda. Además, esta arquitectura destaca por su naturaleza extensible, lo que facilita la integración de nuevas dimensiones contextuales sin comprometer la estructura del sistema. Finalmente, al basarse en un grafo relacional de los POIs, el modelo garantiza una interpretabilidad inherente; esto no solo supera la opacidad de los enfoques de vanguardia, sino que también sienta las bases para la implementación de explicaciones trazables que justifiquen la pertinencia de las recomendaciones.

Para validar nuestra propuesta, hemos implementado dos algoritmos de recomendación —cadenas de Markov y factorización de matrices— e integrado, en ambos casos, nuestro mecanismo de prefiltrado contextual para analizar su rendimiento. El protocolo experimental se ha realizado utilizando el conjunto de datos *Context Trails* para la ciudad de Tokio [8], lo que nos ha permitido trabajar con rutas turísticas reales enriquecidas con información contextual. Como se detallará en la evaluación, hemos aplicado un método de evaluación más restrictivo que los estándares de la literatura: al limitar la validación exclusivamente al siguiente punto de interés visitado por el usuario en la ruta real, obtenemos una medición más rigurosa y fiel del comportamiento del turista. Los resultados preliminares indican que el contexto actúa como un factor determinante en la elección de destinos, aunque subrayamos la necesidad de una evaluación más exhaustiva para corroborar la robustez de estas observaciones iniciales.

El resto del artículo se estructura de la siguiente manera: la Sección 2 detalla la metodología basada en grafos; la Sección 3 expone el diseño experimental junto con los resultados; y la Sección 4 presenta las conclusiones y líneas de trabajo futuro.

2. Recomendación de Rutas Turísticas Basada en Grafos

El sistema de recomendación de rutas turísticas se basa en la representación del conjunto de datos de *Context Trails* [8]. Este conjunto se obtuvo a partir de tres fuentes diferentes: *Semantic Trails* [3], que contiene la base de interacciones entre usuarios y POIs; *Foursquare* [4], para obtener los detalles de los POIs; y

[3] https://figshare.com/articles/dataset/Semantic_Trails_Datasets/7429076
[4] https://location.foursquare.com/developer/

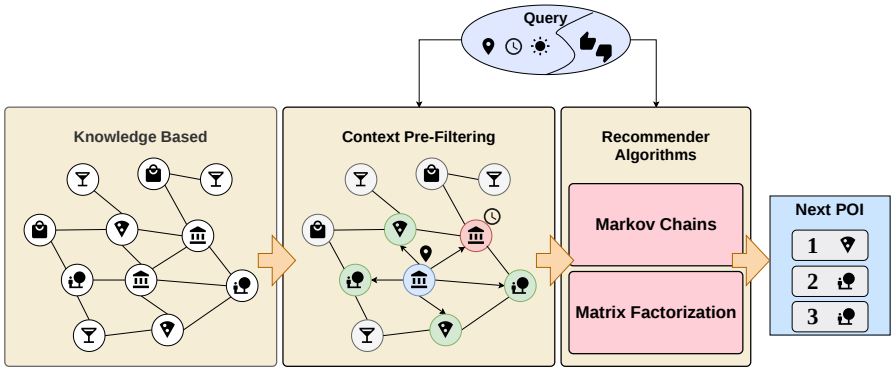


Figura 1. Esquema general del sistema de recomendación. Se estructura en 3 pasos: (1) modelado del dominio y del contexto en un grafo; (2) prefiltrado contextual; y (3) aplicación de modelos de recomendación.

VisualCrossing [5], para obtener datos meteorológicos por horas. Context Trails contiene una descripción detallada de cada POI disponible en cada ciudad, incluyendo las coordenadas geográficas del lugar, su categoría correspondiente dentro de una estructura jerárquica y un horario de apertura definido por 8 intervalos temporales (*early morning*, *morning*, *afternoon* y *night*), diferenciados entre días laborables y fines de semana. Por otro lado, la actividad de los usuarios se representa como una secuencia de interacciones o trayectos, ordenados cronológicamente. Cada una de estas interacciones viene enriquecida tanto con la hora en la que se realizó el desplazamiento como con los atributos climáticos específicos —tales como la condición general del cielo, la temperatura, la velocidad del viento y el nivel de precipitación— registrados en el momento preciso de la llegada al POI de destino.

A partir de esta representación de datos, hemos diseñado un sistema compuesto de tres pasos para la recomendación de rutas turísticas, como se muestra en la Figura 1. Primero, partiendo de la representación de datos que nos ofrece el conjunto de datos de Context Trails, modelamos el conocimiento en un grafo dirigido enriquecido mapeando POIs y transiciones junto a sus dimensiones climáticas y horarias. Segundo, ejecutamos un prefiltrado contextual que reduce el espacio de búsqueda al generar un subgrafo acotado por la consulta. Por último, este subgrafo puede utilizarse en diversos algoritmos de recomendación. En este trabajo, las pruebas se realizaron con un algoritmo basado en cadenas de Markov y otro en la factorización de matrices.

2.1. Modelado del Grafo y Dimensiones Contextuales

El modelo propuesto sitúa al grafo enriquecido como la estructura fundacional para obtener los candidatos en la recomendación de rutas, representando

[5] https://www.visualcrossing.com/

los itinerarios históricos seguidos por los usuarios. En este esquema, cada nodo almacena la información de un punto de interés (POI) e incorpora sus restricciones contextuales: su localización geográfica y su horario de apertura. Por su parte, las aristas orientadas modelan la interacción secuencial de un usuario al desplazarse entre dos puntos. Esta transición integra el contexto dinámico capturado en el instante de llegada al destino, registrando las condiciones climáticas específicas y la hora exacta en la que se completó el trayecto.

Formalmente, estructuramos la base del conocimiento mediante un grafo dirigido $G = (V, E)$, donde $V = \{v_1, v_2, \ldots, v_n\}$ representa el conjunto de POIs, y $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ define el conjunto de aristas orientadas que conectan dichos puntos. Cada vez que un usuario realiza un desplazamiento entre dos POIs, dicho desplazamiento se representa en el grafo mediante una arista. Esta representación matemática resulta idónea para nuestro propósito, ya que permite mapear directamente los recorridos secuenciales y las trayectorias históricas de los usuarios a través de la ciudad. Además, la adopción de este modelado aporta una justificación metodológica crítica, puesto que nos permite capturar la conectividad lógica entre los POIs a partir del comportamiento real de los turistas, evitando así la recomendación de destinos que puedan incumplir las restricciones de localización implícitas en las transiciones físicas cotidianas.

A partir de esta estructura de grafo básica, se procede a modelar e incluir la información contextual tanto de los puntos de interés como de las interacciones secuenciales de los usuarios:

- **Contexto de los nodos:** A cada vértice $v \in V$ se le asocian las restricciones operativas e intrínsecas del POI correspondiente. En primer lugar, se vincula al nodo su localización mediante un par ordenado de coordenadas geográficas $\text{coord}(v) = (\text{lat}_v, \text{lon}_v)$, que representan su latitud y longitud. En segundo lugar, se incluye el horario de apertura como una dimensión contextual estática dentro del nodo. El horario de apertura de un POI se formaliza matemáticamente como un vector de características booleanas $\mathbf{h}_v = [h_1^{\text{wd}}, h_2^{\text{wd}}, h_3^{\text{wd}}, h_4^{\text{wd}}, h_1^{\text{we}}, h_2^{\text{we}}, h_3^{\text{we}}, h_4^{\text{we}}] \in \{0,1\}^8$, donde *wd* representa los días entre semana (*weekday*) y *we* los fines de semana (*weekend*). Por su parte, los subíndices asignan secuencialmente las ventanas temporales discretas del día: *early morning* (1), *morning* (2), *afternoon* (3) y *night* (4).
- **Contexto de las aristas:** Cada arista dirigida $e = (v_i, v_j) \in E$ representa una interacción o un desplazamiento secuencial de un usuario desde un punto de interés de origen v_i hacia uno de destino v_j. A cada una de estas transiciones se le incorporan exclusivamente las dimensiones del contexto dinámico registradas en el instante específico de llegada al POI de destino v_j, omitiendo las condiciones del punto de partida para optimizar el almacenamiento. En primer lugar, se asocia la condición climática en destino mediante una variable categórica $c_e \in C$, donde el conjunto discreto $C = \{c_1, c_2, c_3, c_4, c_5\}$ define los estados meteorológicos posibles, mapeando de forma indexada las condiciones de *cielo despejado* (c_1), *parcialmente nublado* (c_2), *cielo cubierto* (c_3), *lluvia* (c_4) y *nieve* (c_5) extraídas del dominio operativo de los datos. En segundo lugar, se integra el factor temporal de llegada mediante la variable

t_e, que almacena el valor numérico de la hora exacta en la que se completó el trayecto y el usuario interactuó con el vértice de destino.

Este grafo enriquecido constituye la estructura básica de conocimiento sobre la cual se cimienta el modelo de recomendación propuesto, integrando de manera unificada las dimensiones relacionales y contextuales del dominio turístico. Una de las ventajas fundamentales de este diseño semántico es su alta extensibilidad, lo que permite incorporar con facilidad nuevas variables en el futuro sin alterar la arquitectura base. De este modo, tanto los nodos como las aristas pueden expandirse para albergar atributos contextuales adicionales, tales como el tipo de transporte empleado en los desplazamientos entre POIs o el etiquetado específico de rutas y puntos de interés según criterios de sostenibilidad [1].

2.2. Prefiltrado Basado en Contexto

El siguiente paso consiste en aplicar un prefiltrado contextual del grafo para optimizar la eficiencia y la precisión del modelo recomendador, reduciendo el espacio de búsqueda en el momento de la consulta de la recomendación. Este proceso da como resultado un subgrafo con las interacciones históricas y los POIs que se visitaron históricamente en las mismas condiciones contextuales. Esto garantiza que los algoritmos de ordenación posteriores operen exclusivamente sobre un conjunto de candidatos válidos para el usuario en el momento de la recomendación.

Para llevar a cabo este prefiltrado, el sistema procesa una consulta o *query* activa que encapsula el estado del usuario en el momento de la petición. Esta consulta se define como un vector $q = (u, loc, t, c)$, donde u representa el identificador único del usuario, loc denota su localización geoespacial actual, t indica la hora de la consulta y c especifica las restricciones climáticas en dicho instante.

El algoritmo de prefiltrado contextual se ejecuta mediante un proceso secuencial de tres etapas:

1. **Filtrado de Disponibilidad Temporal:** Se evalúa el vector de disponibilidad $\mathbf{h}_v$ de cada POI en la hora t indicada en la *query*. El sistema descarta todos aquellos nodos v cuyo estado en la ventana temporal correspondiente a t sea de inactividad (0), reteniendo exclusivamente los puntos de interés que se encuentran operativos en ese instante.
2. **Filtrado de Interacciones Climáticas:** Sobre el conjunto de aristas E que conectan POIs que se mantienen del paso anterior, se aplica una restricción estricta basada en el clima c. Se seleccionan únicamente aquellas transiciones $e = (v_i, v_j)$ cuyas propiedades contextuales c_e coinciden con la condición climática c incluida en la consulta, eliminando del subgrafo cualquier arista que no sea coherente con el clima actual.
3. **Anclaje Espacial:** El sistema identifica el nodo de referencia $v_{ref} \in V$ que minimiza la distancia geoespacial con respecto a la localización loc de la consulta. Una vez localizada esta ancla, el grafo se restringe únicamente a

los vecinos directos de v_{ref} que han superado los filtros temporales y climáticos previos, delimitando así el conjunto final de candidatos para la fase de recomendación.

Finalmente, el subgrafo resultante constituye el conjunto depurado de candidatos que servirán de base para que los modelos de recomendación generen el ranking final. Dado que este subgrafo preserva la información de las interacciones y de los POIs visitados históricamente bajo condiciones contextuales equivalentes a las de la consulta, la recomendación se realiza mediante una estrategia colaborativa a partir del comportamiento real de los usuarios que han tenido anteriormente [9]. Esta alineación garantiza que el proceso de ordenación no solo opere sobre elementos físicamente factibles y temporalmente viables, sino que se sustente estrictamente en el conocimiento empírico de las rutas recorridas previamente en situaciones similares.

2.3. Recomendación del Siguiente Punto de Interés

Para abordar la generación de recomendaciones, hemos implementado una aproximación híbrida que combina dos paradigmas complementarios: *cadenas de Markov* y *factorización de matrices*. Por un lado, las cadenas de Markov nos permiten enfocar la recomendación en la naturaleza secuencial de los desplazamientos, capturando las rutas históricas que han seguido los usuarios. Por otro lado, la factorización de matrices facilita la integración de las preferencias latentes de los usuarios, al modelar su afinidad intrínseca por determinados POIs.

El valor de predicción final de un POI candidato v, denotado como $\hat{s}(v)$ se obtiene mediante una combinación lineal ponderada de las puntuaciones individuales obtenidas con ambos enfoques. Esta formulación permite equilibrar la importancia relativa entre la probabilidad secuencial y las preferencias latentes del usuario:

$$\hat{s}(v) = \alpha \cdot S_M(v) + (1 - \alpha) \cdot S_F(v) \tag{1}$$

donde $S_M(v)$ representa la puntuación obtenida por el modelo de cadenas de Markov, $S_F(v)$ denota la puntuación derivada de la factorización de matrices, y $\alpha \in [0, 1]$ es un hiperparámetro que regula la influencia de cada componente en la recomendación final. Finalmente, tras obtener las puntuaciones de predicción para cada candidato, el sistema genera un ranking de puntos de interés, ordenando los candidatos de forma descendente según sus valores de $\hat{s}(v)$ para presentar al usuario las opciones más relevantes. A continuación, se detallan ambos modelos.

Modelo de Cadenas de Markov Las cadenas de Markov son modelos estocásticos que representan una secuencia de eventos en los que la probabilidad de transición a un estado futuro depende únicamente del estado actual [7]. Adaptamos este paradigma a nuestro grafo, donde los estados se corresponden con

los POIs y las transiciones representan la probabilidad de que un usuario decida desplazarse hacia un destino específico a partir de su ubicación actual.

La probabilidad de pasar de un POI de origen v_i a un destino v_j se calcula a partir del subgrafo generado tras el prefiltrado, cuantificando la frecuencia relativa de las interacciones que han tenido lugar desde el nodo de origen:

$$P(v_j|v_i) = \frac{count(v_i \rightarrow v_j)}{\sum_{v_k \in V} count(v_i \rightarrow v_k)} \tag{2}$$

Adicionalmente, con el objetivo de dotar al sistema de una mayor capacidad de exploración, aplicamos una técnica de suavizado a las probabilidades calculadas. Esto garantiza que aquellos POIs que no cuentan con una transición explícitamente registrada en el subgrafo actual mantengan una probabilidad mínima de selección, evitando así asignarles una probabilidad nula a destinos que, aunque poco frecuentes en el contexto actual, continúan siendo alternativas plausibles para el usuario. Para el suavizado aplicamos el método de Jelinek-Mercer [10].

Finalmente, el valor resultante del suavizado representa el *score* de verosimilitud de cada candidato. Este valor cuantifica la idoneidad del destino considerando tanto la evidencia contextual, la frecuencia de transición bajo condiciones similares, como la evidencia histórica, la frecuencia global de la ruta. En el proceso de generación de recomendaciones, estas probabilidades sirven como base para ordenar los POIs candidatos, permitiendo al sistema priorizar aquellas transiciones que, según el historial del grafo, poseen una mayor coherencia secuencial y viabilidad contextual con respecto al estado actual del usuario.

Factorización de Matrices La Factorización de Matrices (MF) tiene como objetivo modelar y representar las preferencias intrínsecas de los usuarios al consumir puntos de interés, capturando la afinidad latente entre cada usuario y cada POI. La técnica de factorización de matrices descompone la matriz de interacción original en dos matrices de rango bajo, proyectando tanto a los usuarios como a los POIs en un espacio vectorial compartido de dimensiones reducidas [2]. Esto permite estimar la preferencia de un usuario por un POI, incluso cuando no existe una interacción histórica directa entre ambos. Para este trabajo hemos usado la propuesta de *Non-Negative Matrix Factorization* [5].

A diferencia del modelo de Markov, este algoritmo se entrena con el conjunto completo de interacciones registradas en la base de datos, omitiendo explícitamente cualquier dimensión contextual. De este modo, el modelo de factorización de matrices aprende una representación de los gustos estáticos y globales del usuario. Para integrar este enfoque en nuestro sistema, tomamos los POIs candidatos presentes en el subgrafo generado tras el prefiltrado contextual y los reordenamos según las puntuaciones de preferencia estimadas por el modelo, permitiendo que las afinidades personales del usuario influyan en la selección final dentro del entorno operativo restringido.

En este sentido, el valor $S_F(v)$ representa la puntuación de afinidad estimada del usuario u respecto al POI v, calculada habitualmente mediante el producto escalar entre sus respectivos vectores latentes. Esta puntuación captura el grado

de preferencia global del usuario, proporcionando una métrica complementaria a la probabilidad de transición local. Por tanto, mientras que el componente de Markov prioriza la coherencia secuencial de la ruta dentro del subgrafo, el valor $S_F(v)$ actúa como un mecanismo de personalización que pondera aquellos POIs que mejor se alinean con el perfil histórico del usuario, permitiendo que las afinidades personales influyan en la selección final dentro del entorno operativo restringido.

3. Evaluación y Resultados

En esta sección presentamos una evaluación preliminar de la propuesta metodológica detallada en los apartados anteriores. El objetivo fundamental de este análisis es determinar la precisión y la eficacia de nuestro enfoque para recomendar rutas turísticas completas. Asimismo, buscamos estudiar el impacto real de las variables contextuales, como la meteorología y la disponibilidad temporal, en las decisiones de navegación de los usuarios, validando así la capacidad del sistema para adaptar sus recomendaciones a situaciones específicas. Para abordar este análisis, hemos estructurado la sección en dos subapartados: en primer lugar, se describe la configuración experimental, incluyendo la metodología de particionado de datos y las métricas empleadas; a continuación, se exponen y discuten los resultados obtenidos, analizando el rendimiento del modelo bajo distintas configuraciones.

3.1. Configuración de la evaluación

Para validar la eficacia de nuestra propuesta, hemos diseñado un protocolo experimental exhaustivo que compara la precisión de nuestro modelo híbrido frente a diversas líneas base aplicando el subgrafo obtenido tras el prefiltrado contextual.

Datos utilizados Para este estudio experimental, hemos utilizado el conjunto de datos *Context Trails* [8], que recopila trayectorias turísticas reales geoetiquetadas. Específicamente, nos hemos centrado en los datos de la ciudad de Tokio, seleccionada por su alta densidad de POIs y la gran variabilidad de las condiciones climáticas y temporales que presenta, lo que permite evaluar adecuadamente la robustez de nuestro modelo frente a dicho contexto. Para este conjunto de datos de la ciudad de Tokio, contamos con 64.087 POIs, 586.949 interacciones y 47.331 rutas turísticas.

División de los datos Para asegurar la fiabilidad de los resultados, hemos aplicado una estrategia de particionado temporal. Los datos han sido divididos en dos conjuntos: un 80 % destinado al entrenamiento de los modelos y un 20 % reservado para el conjunto de test. Esta división se ha realizado manteniendo la integridad cronológica de las rutas a nivel de usuario. Es decir, el 20 % de

las rutas más modernas de cada usuario se han usado en test. Esto asegura que el sistema sea evaluado sobre trayectorias que ocurrieron posteriormente a las utilizadas durante su fase de aprendizaje, simulando así una situación real de recomendación.

Modelos de comparación Para evaluar el rendimiento de nuestra propuesta híbrida, hemos seleccionado dos modelos de referencia como líneas base elementales que actúan como puntos de comparación:

1. **Random:** Un modelo que selecciona candidatos de forma puramente aleatoria entre los POIs disponibles.
2. **MostPop:** Un modelo que recomienda POIs basándose exclusivamente en su frecuencia global de visitas en los datos, sin considerar el historial del usuario ni el contexto.

La evaluación experimental se realiza comparando estos dos modelos con las soluciones propuestas en este trabajo: el modelo secuencial basado en cadenas de Markov, el modelo de preferencias mediante Factorización de Matrices (MF) y el enfoque híbrido bajo distintas configuraciones de pesos. De este modo, se busca cuantificar con precisión la mejora metodológica que aportan tanto la modelización de las transiciones secuenciales como el aprendizaje de las preferencias latentes del usuario frente a los enfoques tradicionales.

Protocolo de evaluación El proceso de evaluación está diseñado para medir el rendimiento de los algoritmos exclusivamente a partir del subgrado resultante tras el prefiltrado contextual, en el que solo se consideran los candidatos coherentes con el momento del día y el clima en el momento de la consulta.

En cuanto a la metodología de medición, nuestro proceso sigue un enfoque de recomendación secuencial paso a paso (siguiente punto de interés). Para cada ruta del conjunto de test, el sistema determina la ubicación del usuario utilizando el primer POI de la secuencia, mientras que las variables contextuales de clima y del intervalo del tiempo se extraen directamente del instante en el que el usuario visitó el siguiente POI de su trayectoria real. Con esta información, el sistema ejecuta los modelos de recomendación y genera un ranking de candidatos, evaluando el éxito del algoritmo según si el POI real visitado por el usuario figura en la lista recomendada. Para evaluar el siguiente punto de la ruta, volvemos a coger como referencia el POI de test que ha realizado realmente el usuario para volver a calcular el siguiente punto de la ruta. Este proceso se repite hasta haber completado el número de POIs que ha visitado el usuario en su ruta. Cabe destacar que este protocolo de evaluación es considerablemente más exigente que otros métodos convencionales; si bien este enfoque puede arrojar valores numéricos más conservadores en las métricas de rendimiento, proporciona una medida significativamente más realista, al validar la capacidad del modelo para predecir con precisión las decisiones de navegación del usuario en rutas completas bajo condiciones contextuales reales.

Tabla 1. Comparativa de la precisión de los modelos de recomendación. En negrita se muestra el mejor valor de cada métrica; el segundo mejor valor está subrayado.

	HR@K			MRR@K		NDCG@K	
Modelo	**1**	**5**	**10**	**5**	**10**	**5**	**10**
Random	0.0072	0.0224	0.0335	0.0124	0.0138	0.0149	0.0184
Popularity	<u>0.0119</u>	0.0347	0.0484	0.0198	0.0216	0.0234	0.0279
MF	0.0070	0.0225	0.0330	0.0122	0.0136	0.0147	0.0181
Markov	**0.0125**	**0.0396**	**0.0592**	**0.0219**	**0.0244**	**0.0262**	**0.0325**
Markov+MF (20-80)	0.0090	0.0289	0.0450	0.0158	0.0179	0.0191	0.0242
Markov+MF (50-50)	0.0110	0.0335	0.0515	0.0187	0.0211	0.0224	0.0282
Markov+MF (80-20)	<u>0.0119</u>	<u>0.0382</u>	<u>0.0566</u>	<u>0.0210</u>	<u>0.0234</u>	<u>0.0252</u>	<u>0.0312</u>

Para poder evaluar la precisión tanto de los baselines como del modelo propuesto, se han aplicado métricas que miden la calidad de los rankings generados por cada uno de los pasos de la ruta:

- **Hit Rate (HR@K):** Mide la proporción de casos en los que el POI real visitado aparece entre los K primeros resultados recomendados.
- **Mean Reciprocal Rank (MRR@K):** Evalúa la calidad del ranking midiendo la posición inversa del primer POI relevante encontrado, penalizando las recomendaciones en las que el acierto aparece en posiciones inferiores.
- **Normalized Discounted Cumulative Gain (NDCG@K):** Mide la calidad del ranking considerando tanto la relevancia del ítem como su posición en la lista recomendada, otorgando un mayor peso a los aciertos situados en las posiciones superiores del ranking.

3.2. Resultados

A continuación, se presentan los resultados experimentales obtenidos tras evaluar los distintos modelos de recomendación sobre el conjunto de datos de Tokio. La Tabla 1 detalla los valores de precisión de los métodos propuestos (MF y Markov) frente a las líneas base (Random y Popularity), todos ellos operando bajo el mismo escenario de prefiltrado contextual.

Como se observa, el modelo basado en cadenas de Markov muestra una precisión superior en todas las métricas evaluadas. Dado que todos los algoritmos compiten bajo las mismas restricciones, utilizando únicamente POIs abiertos en el horario de consulta y transiciones realizadas por otros usuarios bajo las mismas condiciones meteorológicas, este resultado confirma que la dependencia secuencial entre los POIs es el factor determinante en la navegación turística, superando claramente a los enfoques basados en preferencias estáticas, como la factorización de matrices. La representación de los datos en un grafo junto con el prefiltrado del mismo permite capturar la dinámica de desplazamiento con mayor fidelidad al integrar las variables contextuales directamente en la arquitectura del sistema. Si bien estos resultados iniciales son prometedores, es preciso

ampliar el alcance de la evaluación mediante el análisis en otros entornos geográficos y realizar comparativas con propuestas del estado del arte para consolidar y validar estas conclusiones.

Por otro lado, el modelo de Factorización de Matrices (MF) muestra un rendimiento inferior al de la línea base Popularity, lo que además provoca que el modelo híbrido pierda precisión a medida que aumenta el peso de MF. Este fenómeno se atribuye al marcado sesgo de popularidad del conjunto de datos y a la naturaleza altamente restrictiva del prefiltrado contextual. Al reducir drásticamente los POIs candidatos según el clima y el horario, el espacio de búsqueda resultante puede comprometer la satisfacción del usuario al ignorar sus preferencias latentes a largo plazo. Como línea de trabajo futura, resulta prometedor explorar mecanismos de prefiltrado menos agresivos o enfoques híbridos que equilibren mejor las restricciones del contexto con los gustos individuales, o que minimicen el sesgo de popularidad de los POIs.

En resumen, los resultados iniciales demuestran que integrar el contexto directamente en las aristas del grafo aporta eficiencia e interpretabilidad, y que la dispersión de datos derivada del prefiltrado constituye su principal limitación. Para solventar esto y validar la robustez general del modelo, el trabajo futuro se centrará en ampliar la evaluación mediante la aplicación de nuevos algoritmos y la prueba del sistema con conjuntos de datos de otras ciudades y entornos urbanos.

4. Conclusiones y Trabajo Futuro

En este trabajo hemos abordado el desafío de recomendar rutas turísticas sensibles al contexto. Nuestra propuesta introduce una arquitectura basada en grafos que modela los Puntos de Interés (POIs) como nodos y las transiciones de los usuarios como aristas enriquecidas con información meteorológica y temporal. Mediante un mecanismo de prefiltrado contextual, optimizamos el espacio de búsqueda, mejorando la eficiencia y la interpretabilidad del sistema frente a modelos de caja negra. La evaluación realizada, bajo un protocolo estricto de predicción del siguiente punto de interés, demuestra que la dependencia secuencial es crítica en la navegación turística, proporcionando una base sólida para futuras extensiones.

A pesar de asegurar la viabilidad operativa, el enfoque actual presenta limitaciones en la cobertura por dispersión de datos debido a un prefiltrado contextual excesivamente rígido, sumado a un sesgo de popularidad heredado de la concentración de rutas históricas. Para abordar estos desafíos, las líneas de trabajo futuras se orientarán a flexibilizar el filtrado, mitigar el sesgo y enriquecer el grafo con nuevas dimensiones, como los modos de transporte y la sostenibilidad. Finalmente, se aprovechará la naturaleza del grafo para generar recomendaciones explicables y se contrastará el rendimiento del sistema frente a los algoritmos de referencia de la literatura.

Acknowledgments. Esta investigación ha sido financiada por los proyectos PID2022-139131NB-100 y PID2025-167968NB-I00, financiados por el MCIN/AEI/10.13039/501-

100011033 y el FEDER "Una manera de hacer Europa". También cuenta con la financiación del proyecto SI4/PJI/2024-00032 financiado por la Comunidad de Madrid a través del convenio-subvención, para el fomento y la promoción de la investigación y la transferencia de tecnología en la Universidad Autónoma de Madrid.

Referencias

1. Banerjee, A., Mahmudov, T., Adler, E., Aisyah, F.N., Wörndl, W.: Modeling Sustainable City Trips: Integrating CO2e Emissions, Popularity, and Seasonality into Tourism Recommender Systems. Information Technology & Tourism **27**(1), 189–226 (2025). https://doi.org/10.1007/s40558-024-00303-1
2. Koren, Y., Rendle, S., Bell, R.: Advances in Collaborative Filtering. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 91–142. Springer US, New York, NY (2022). https://doi.org/10.1007/978-1-0716-2197-4_3
3. Lei, Y., Shen, L., Sun, Z., He, T.: Context-adaptive Graph Neural Networks for Next POI recommendation. Electronic Commerce Research and Applications **77**, 101597 (2026). https://doi.org/10.1016/j.elerap.2026.101597
4. Lu, Y.S., Huang, J.L.: GLR: A graph-based latent representation model for successive POI recommendation. Future Generation Computer Systems **102**, 230–244 (2020). https://doi.org/10.1016/j.future.2019.07.074
5. Luo, X., Zhou, M., Xia, Y., Zhu, Q.: An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems. IEEE Transactions on Industrial Informatics **10**(2), 1273–1284 (2014). https://doi.org/10.1109/TII.2014.2308433
6. Rahmani, H.A., Aliannejadi, M., Baratchi, M., Crestani, F.: A Systematic Analysis on the Impact of Contextual Information on Point-of-Interest Recommendation. ACM Transactions on Information Systems (TOIS) **40**(4), 88:1–88:35 (2022). https://doi.org/10.1145/3508478
7. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized Markov chains for next-basket recommendation. pp. 811–820. WWW '10 (2010). https://doi.org/10.1145/1772690.1772773
8. Sánchez, P., Bellogin, A., Jorro-Aragoneses, J.L.: Context Trails: A Dataset to Study Contextual and Route Recommendation. pp. 716–725. RecSys '25 (2025). https://doi.org/10.1145/3705328.3748151
9. Tran, T., Phung, D., Venkatesh, S.: Collaborative filtering via sparse Markov random fields. Information Sciences **369**, 221–237 (2016). https://doi.org/10.1016/j.ins.2016.06.027
10. Valcarce, D., Parapar, J., Barreiro, Á.: A Study of Smoothing Methods for Relevance-Based Language Modelling of Recommender Systems. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds.) Advances in Information Retrieval. pp. 346–351 (2015). https://doi.org/10.1007/978-3-319-16354-3_38
11. Xing, S., Liu, F., Wang, Q., Zhao, X., Li, T.: Content-aware point-of-interest recommendation based on convolutional neural network. Applied Intelligence **49**(3), 858–871 (2019). https://doi.org/10.1007/s10489-018-1276-1
12. Yao, L., Sheng, Q.Z., Qin, Y., Wang, X., Shemshadi, A., He, Q.: Context-aware Point-of-Interest Recommendation Using Tensor Factorization with Social Regularization. pp. 1007–1010. SIGIR '15 (2015). https://doi.org/10.1145/2766462.2767794