

Sistema de recomendación híbrido

Ana Paula González Muñoz
Dennis Daniel González Durán

Universidad de La Habana, Cuba

Resumen. En este trabajo, se presenta un sistema de recomendación híbrido diseñado para sugerir destinos turísticos personalizados en ciudades desconocidas para los usuarios. El sistema integra diversos enfoques, combinando técnicas de análisis de grafos, reglas basadas en conocimiento y métodos colaborativos para superar las limitaciones de los sistemas de recomendación tradicionales. El modelo propuesto unifica la relación entre usuarios y ubicaciones, usuarios entre sí, y entre las propias ubicaciones en un marco que mejora relevancia de las recomendaciones. Se emplea el método de Random Walk with Restart (RWR) para optimizar las recomendaciones considerando las interacciones complejas dentro de un grafo ponderado. Los experimentos realizados con un conjunto de datos del mundo real demuestran la capacidad del sistema para ofrecer recomendaciones valiosas, destacando su eficacia a pesar de la baja precisión en algunos casos. El enfoque equilibrado entre simplicidad y funcionalidad del sistema lo convierte en una herramienta útil y accesible para la planificación de viajes.

Palabras clave: Sistema de recomendación híbrido · Random Walk with Restart · Recomendación de viajes · Análisis de grafos · Personalización de recomendaciones · Sistema basado en conocimiento.

1 Introducción

En el contexto actual de la personalización digital, los sistemas de recomendación híbridos combinan diversas técnicas para ofrecer recomendaciones más precisas y relevantes. La construcción de sistemas híbridos[1] que integren las fortalezas de diferentes algoritmos y modelos se ha convertido en el objetivo de investigaciones recientes para superar algunas de las deficiencias y problemas de los métodos individuales. Estos sistemas, aunque más complejos y exigentes en términos de recursos computacionales, presentan oportunidades significativas para mejorar la calidad de las recomendaciones.

1.1 Actualidad y sistema propuesto

La recomendación personalizada de viajes ha tenido diversos enfoques en su investigación, destacando sistemas como Cyberguide[2] y TripTip[3], que utilizan la ubicación del usuario, su historial y similitudes entre lugares visitados para sugerir destinos. Otros estudios han explorado datos generados por usuarios,

como fotos geoetiquetadas[4] y check-ins[5][6], para mejorar las recomendaciones. Sin embargo, muchas soluciones dependen de información de expertos, que puede ser limitada y desactualizada.

Se presenta un sistema de recomendación híbrido de viajes personalizado que aborda algunas de las deficiencias anteriores y recomienda a los usuarios dónde ir en una ciudad desconocida. Construye un modelo que unifica la relación de ubicación de usuario, la relación usuario-usuario y la relación ubicación-ubicación en un solo marco. Además para realizar la recomendación se tiene en cuenta las preferencias de los usuarios y las características de las locaciones.

2 Preliminares

En esta sección, se presentan los conceptos fundamentales y las bases teóricas necesarias para comprender el enfoque propuesto en este trabajo.

2.1 Formulación del problema

En los sistemas de recomendación[1] se denota el conjunto de usuarios como $U = \{u_1, u_2, \dots, u_m\}$ y el conjunto de locaciones como $V = \{v_1, v_2, \dots, v_n\}$, donde $m = |U|$ y $n = |V|$. El objetivo de este sistema es predecir los lugares que un usuario debería visitar, en una ciudad en específico, a partir de un registro de valoraciones de una comunidad de usuarios a estos lugares y las características de los mismos.

2.2 Recomendación basada en conocimiento

Un sistema de recomendación basado en conocimiento[1] es una tecnología diseñada para sugerir productos, servicios o información a los usuarios en función de su conocimiento explícito sobre sus preferencias, necesidades y características. Este tipo de sistema utiliza un conjunto estructurado de reglas y criterios que se alimentan de información detallada sobre los productos o servicios y las necesidades específicas del usuario.

Primeramente, el sistema posee un conjunto de reglas simples para incrementar la relevancia de una localidad para un usuario, por ejemplo:

- si el usuario visita con frecuencia localidades de cierta categoría, aumentar la relevancia de esta en la selección de localidades.
- si el usuario prefiere lugares de alto o bajo presupuesto, recomendar localidades en consecuencia.
- si el usuario pertenece a cierto grupo de edad, recomendarle lugares de categoría con buenas valoraciones de ese grupo.

2.3 *Random Walk with Restart*

El método Random Walk with Restart (**RWR**)[7] se utiliza en diversas áreas, como la recomendación social y de lugares, al calcular la relevancia entre nodos en un grafo ponderado. **RWR** captura la estructura global del grafo y las relaciones entre nodos. Se destacan estudios que usan RWR para recomendar música[8] y lugares[9], aunque con limitaciones como no considerar todas las relaciones posibles.

3 Arquitectura del sistema

El sistema de recomendación utiliza un enfoque híbrido combinando reglas sencillas y análisis de grafos para personalizar las recomendaciones de localidades. Primero, el sistema aplica reglas basadas en las características de los usuarios y sus valoraciones para identificar localidades de interés. Luego, construye un grafo que representa las relaciones entre usuarios y localidades, así como entre los usuarios y las localidades entre sí (modelo 3R [10]). Este grafo se emplea para enriquecer el análisis de las recomendaciones empleando RWR, capturando interacciones complejas y patrones adicionales. Finalmente, se generan recomendaciones personalizadas para cada usuario, integrando tanto las sugerencias basadas en reglas como el análisis del grafo para ofrecer opciones más precisas y relevantes.

4 Marco de trabajo

En esta sección, se detallarán los pasos necesarios para la implementación del sistema de recomendación propuesto.

4.1 Extracción de datos

Se trabaja con un conjunto de datos que contiene varias atracciones turísticas en 5 ciudades principales de Indonesia ¹. Este conjunto de datos también consta de 4 archivos:

- *tourism_with_id.csv*: contiene información sobre atracciones turísticas en 5 ciudades principales de Indonesia, con un total de aproximadamente 400.
- *user.csv*: contiene datos ficticios de usuarios para crear funciones de recomendación.
- *tourism_rating.csv*: contiene 3 columnas: el usuario, el lugar y la calificación otorgada; sirve para crear un sistema de recomendación basado en las calificaciones.
- *package_tourism.csv*: contiene recomendaciones de lugares cercanos basados en tiempo, costo y calificación.

¹ <https://www.kaggle.com/datasets/aprabowo/indonesia-tourism-destination/data>

Del conjunto de datos de entrada se extrae información como edad de los usuarios, preferencias de los grupos de edades con respecto a la categoría y precio de los lugares a visitar. También se extraen las preferencias de los usuarios en base a las valoraciones emitidas por estos.

Para obtener la frecuencia con que un usuario visita atracciones de cierta categoría (**u_freq**), se calcula proporción entre atracciones visitadas de esa categoría por el usuario y el total de atracciones visitadas por el mismo.

Para obtener el promedio de dinero que emplea el usuario en atracciones de una categoría (**u_prec**), se calcula el promedio de los precios de las atracciones de esta categoría visitadas por el usuario.

Para obtener el promedio de valoración que emite el usuario en atracciones de una categoría (**u_val**), se calcula el promedio de las valoraciones de las atracciones de esta categoría visitadas por el usuario.

Análogamente, se realizan las mismas extracciones de datos para los grupos de edades: **e_freq**, **e_prec**, **e_val**.

4.2 Manejo del conocimiento

En esta etapa se emplean reglas predefinidas para calcular la relevancia de las localidades en la recomendación del usuario. Se utilizan los datos extraídos en la sección 4.1 para generar una primera escala de los lugares más importantes para el usuario.

La relevancia de una localidad para un usuario se calcula:

$$rel_{u,v} = u_freq_v * \frac{1}{|u_prec_v - prec_v|} * u_val_v * eu_freq_v * \frac{1}{|eu_prec_v - prec_v|} * eu_val_v$$

donde: u_freq_v representa el calculo de **u_freq** para la categoría de v , u_prec_v representa el calculo de **u_prec** para la categoría de v , $prec_v$ representa el precio de v , u_val_v representa el calculo **u_val** para la categoría de v , eu_freq_v representa el calculo de **e_freq** del rango de edad de u para la categoría de v , eu_prec_v representa el calculo de **e_prec** del rango de edad de u para la categoría de v , eu_val_v representa el calculo de **e_val** del rango de edad de u para la categoría de v .

4.3 Análisis colaborativo

En esta etapa se construye un modelo que refleja 3 tipos de relaciones: usuario-usuario, usuario-localidad, localidad-localidad usando un enfoque colaborativo. [10][11]

Relación Usuario-Usuario: se define como las relaciones sociales entre usuarios y se ha utilizado tempranamente en muchos recomendadores. Se calcula como la similitud de coseno entre el vector de valoraciones de dos usuarios[11].

Relación Usuario-Localidad: es una característica indispensable utilizada en todos los sistemas de recomendación. En el sistema, en lugar de usar calificaciones simples directamente, se combinan con la popularidad del artículo para determinar el peso entre el usuario y el artículo. Los clientes se dejan influenciar

fácilmente por los comportamientos de los demás. Por lo tanto, se consiera que, los usuarios son más propensos a pasar al artículo con mayor popularidad, y el peso entre usuario y artículo se puede definir como:

$$rel_{u,v} = f_{rating}(u, v) + f_{review}(v)$$

donde: $f_{rating}(u, v)$ representa la valoración normalizada del usuario u a la localidad v y $f_{review}(v)$ representa la popularidad normalizada de la localidad v .

$$f_{rating}(u, v) = \frac{1}{2} + \frac{r_{u,v} - Avg(U_u)}{max(R)}$$

donde: $r_{u,v}$ es la valoración del usuario u a la localidad v , $Avg(U_u)$ es el promedio de las valoraciones emitidas por el usuario y $max(R)$ la valoración máxima que permite el sistema.

$$f_{review}(v) = \frac{1}{2} + \frac{v_u - Avg(V_u)}{max(V)}$$

donde: v_u es el número de valoraciones de la localidad v , $Avg(V_u)$ es el promedio de la cantidad de valoraciones de las localidades y $max(V)$ la máxima cantidad de valoraciones para las localidades.

Relación Localidad-Localidad: muchas investigaciones existentes simplemente establecen la relación localidad-localidad con el valor de similitud, mientras que el principio básico de un modelo cadena de Markov es que un usuario tiene cierta posibilidad de transitar desde/hacia cualquier nodo vecino en un grafo. Es decir, si un usuario llega al elemento i , entonces la probabilidad con la que viaja al elemento j debe equilibrarse con la similitud entre V_i e V_j , y la popularidad actual del V_j . Se calcula como la similitud de coseno entre el vector de valoraciones de dos localidades sumado a la popularidad de la localidad representada por $f_{review}(v)$.

Luego de calcular estas relaciones se procede a construir un grafo representado por una matriz de adyacencia con pesos de dimensión $|U| + |V|$ que representa la ponderación de las aristas de cualquier par de nodos, quedando plasmadas los 3 tipos de relaciones antes descritos. Se procede a normalizar esta matriz, para que la matriz en la posición $[x,y]$ represente la probabilidad de transitar desde el nodo x hacia el nodo y en una iteración.

4.4 RWR

Después de construir el modelo 3R, el problema de clasificar las ubicaciones para un usuario específico se reduce al problema de la evaluación de proximidad entre el nodo de usuario y los nodos de ubicación. RWR proporciona una buena puntuación de relevancia entre dos nodos en un grafo ponderado y, por lo tanto, es muy apropiado para ser utilizado para analizar el modelo. RWR se define como:

$$r_{i+1} = c * M * r_i + (1 - c) * E$$

donde: $(1-c)$ es la probabilidad de regresar al nodo inicial, M es la matriz resultante del análisis colaborativo y E es un vector de tamaño $|U| + |V|$ con todas sus componentes 0 a excepción de la componete inicial que es 1, r_i es un vector de tamaño $|U| + |V|$ donde en la posición x denota la probabilidad de transicionar del nodo actual al nodo x en la iteración i -ésima del RWR. [10]

El recorrido consiste en apoyarse en esta fórmula para en una iteración para decidir a que nodo moverse y generar un nuevo vector r_{i+1} para apoyar a la selección del nodo en la iteración siguiente. Finalmente, al terminar las iteraciones podemos utilizar el vector r_i final para ordenar todos los nodos visitados que sean localidades y devolver al usuario los k más recomendados.

La intuición de realizar RWR en el modelo 3R es que para un viajero específico u , primero seguirá a los otros viajeros que comparten una alta similitud de usuario con ella a algunas atracciones turísticas con una alta probabilidad. Esta es la intuición del filtrado colaborativo basado en el usuario. Luego, para cada atracción turística l , también visitará otras atracciones turísticas que son similares a l con una alta probabilidad. Esta es la intuición del filtrado colaborativo basado en elementos. En otras palabras, u realiza un recorrido aleatorio en el modelo 3R. La ventaja de este método es que podemos encontrar tanto las ubicaciones populares como las ubicaciones relevantes para el viajero. Las primeras ubicaciones se pueden encontrar ya que generalmente hay muchos enlaces vinculados a ellas, mientras que las últimas ubicaciones se pueden encontrar ya que están vinculadas por los viajeros que son similares al viajero específico o por las otras ubicaciones relevantes. En resumen, el modelo 3R puede tener en cuenta la popularidad de la ubicación, la similitud del usuario y la similitud de la ubicación simultáneamente.

5 Experimentos

Para la evaluación, se utiliza el enfoque estándar para dividir el conjunto de datos en dos partes: las ciudades de entrenamiento y la ciudad de prueba. Se seleccionan 100 usuarios que hayan visitado la ciudad de prueba y al menos otra ciudad de entrenamiento como usuarios de prueba. Solo se tienen en cuenta a los usuarios que han visitado al menos 5 ubicaciones en las ciudades de entrenamiento para asegurar que los usuarios de prueba hayan proporcionado una cantidad decente de información de preferencias. A continuación, se filtran las atracciones turísticas que los usuarios de prueba han visitado en la ciudad de prueba (lo que hace que parezca que los usuarios de prueba nunca han estado en la ciudad de prueba).

5.1 Métricas

Para evaluar la calidad de la recomendación, es importante averiguar cuántas atracciones turísticas realmente visitadas por el usuario en los datos de prueba son descubiertas por el algoritmo de recomendación. Para ello, se emplearon métricas como:

- **Precision:** mide la proporción de elementos relevantes entre los N elementos recomendados.
- **Recall:** mide la proporción de elementos relevantes que fueron recomendados entre todos los elementos relevantes disponibles.

- **Hit Ratio (HR)**: mide la proporción de usuarios para los cuales al menos uno de los elementos relevantes aparece en el top-N de las recomendaciones.
- **Normalized Discounted Cumulative Gain (nDCG)**: mide la ganancia acumulada de un conjunto de recomendaciones, normalizada por el mejor orden posible.
- **Mean Reciprocal Rank (MRR)**: mide qué tan rápido un sistema puede recomendar un elemento relevante.

Se utilizó un truncamiento en 10 como métrica de evaluación, es decir *Precision@10*, *Recall@10*, *HR@10*, *nDCG@10* y *MRR@10*.

Table 1. Resultados obtenidos

Precision@10	Recall@10	nDCG@10	MRR@10	HR@10
0.07544	0.33893	0.5893	0.1994	0.2008

5.2 Valoración de los resultados de los experimentos

La baja precisión, con un valor de 0.07544, indica que solo un pequeño porcentaje de los primeros 10 resultados son relevantes. Sin embargo, el sistema muestra una capacidad razonable para recuperar resultados relevantes, con un recall@10 de 0.3394. Esto sugiere que, a pesar de la baja precisión, el sistema logra identificar un número significativo de resultados relevantes en los primeros 10 resultados. El Hit Rate@10 de 0.5894 apoya esta observación, mostrando que el sistema devuelve al menos un resultado relevante en alrededor del 59% de las consultas, lo que es positivo.

La MRR@10, con un valor de 0.1995, y el nDCG@10 de 0.2008, indican que, aunque la calidad en el ranking de los resultados no es la mejor, el sistema mantiene una capacidad decente para priorizar resultados relevantes en las primeras posiciones. En general, aunque la precisión es un área de preocupación, el sistema tiene un desempeño razonable en términos de recuperación y clasificación de resultados relevantes.

Teniendo en cuenta que el sistema es ligero y fácil de entender, evitando métodos complejos o no transparentes, estas métricas reflejan un equilibrio entre simplicidad y funcionalidad. La transparencia y la accesibilidad del sistema facilitan su uso y comprensión, lo que añade valor a pesar de las limitaciones en precisión.

6 Conclusiones

El sistema de recomendación híbrido propuesto demuestra ser efectivo para personalizar recomendaciones de viajes, combinando reglas basadas en conocimiento con análisis de grafos y métodos colaborativos. Aunque la precisión del sistema

puede mejorar, ofrece un buen rendimiento en la recuperación de elementos relevantes, destacando su capacidad para generar sugerencias útiles. El enfoque equilibrado entre simplicidad y funcionalidad hace que el sistema sea accesible y fácil de usar, lo que lo convierte en una herramienta valiosa para la planificación de viajes en contextos reales.

References

1. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. Cambridge University Press, New York (2011)
2. Abowd, G. D., Atkeson, C. G., Hong, J. I., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: A mobile context-aware tour guide. *Wireless Networks* **3**(5), 421–433 (1997)
3. Kim, J., Kim, H., Ryu, J. Hee: Triptip: a trip planning service with tag-based recommendation. In: CHI Extended Abstracts, pp. 3467–3472. ACM, New York (2009)
4. Clements, M., Serdyukov, P., de Vries, A. P., Reinders, M. J. T.: Personalised travel recommendation based on location co-occurrence. *CoRR* **abs/1106.5213** (2011)
5. Ye, M., Yin, P., Lee, W.-C.: Location recommendation for location-based social networks. In: GIS, pp. 458–461. ACM, New York (2010)
6. Ye, M., Yin, P., Lee, W.-C., Lee, D. L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR, pp. 325–334. ACM, New York (2011)
7. Lovász, L.: Random walks on graphs: A survey. In: Combinatorics, Paul Erdős is Eighty, vol. 2, pp. 353–397. József Határ, Budapest (1996)
8. Konstas, I., Stathopoulos, V., Jose, J. M.: On social networks and collaborative recommendation. In: SIGIR, pp. 195–202. ACM, New York (2009)
9. Noulas, A., Scellato, S., Lathia, N., Mascolo, C.: A random walk around the city: New venue recommendation in location-based social networks. In: SocialCom '12, pp. 144–153. IEEE (2012)
10. Guo, L., Shao, J., Tan, K.-L., Yang, Y.: WhereToGo: Personalized Travel Recommendation for Individuals and Groups. In: 15th International Conference on Mobile Data Management, pp. 49–58. IEEE (2014). <https://doi.org/10.1109/MDM.2014.12>
11. Chen, Z., Hu, A., Xu, J., Liu, C. H.: DineTogether: A Social-Aware Group Sequential Recommender System. In: TRIDENTCOM 2017, Dalian, People's Republic of China. EAI, 2018. <https://doi.org/10.4108/eai.8-1-2018.155564>