Ablation Testing for Enhanced Multistakeholder Recommender Systems in Urban Tourism

Ofir Ben David
obenda10@campus.haifa.ac.il
Haifa University
Haifa, Israel

Tamar Kohan
tkohan@campus.haifa.ac.il
Haifa University
Haifa, Israel

Abstract

Urban tourism currently faces a significant challenge known as "overtourism," where popular Points of Interest (POIs) become overcrowded. Through a Multistakeholder Recommender System (MRS), this study addresses the conflict between destination managers, who aim to promote sustainability by distributing visitors to peripheral areas, and tourists, who typically seek recommendations aligned with their preferences, often biased toward popular, crowded sites.

Building upon Merinov & Ricci (2024) [1] simulation framework that demonstrated positive-sum outcomes in Rome, Florence, and Istanbul, this research focuses exclusively on Rome to systematically evaluate five improvement directions through ablation testing. Our approach models users' limited city awareness and stochastic decision-making processes using real Flickr data.

We implemented and evaluated five improvements:

(1) User-centric modeling of preferred distance from city center as a continuous variable rather than binary classification (city center vs. periphery).
(2) K-Nearest Neighbors models to better handle sparse tourism data compared to matrix factorization and autoencoders approaches.
(3) Triple-stakeholder framework incorporating religious/spiritual site promotion alongside user satisfaction and sustainability goals.
(4) POI system enhancements including reduced walking logic and smart routing optimization.
(5) LLM-enhanced semantic initialization using POI descriptions to improve cold-start handling.

Results from 615 simulations across varying parameters (user history size, city awareness levels, and trade-off parameters) demonstrate significant improvements. The POI system enhancement with smart routing achieved the most substantial gains (+253% in quantity metrics and +436% in quality metrics, with a 54% walking distance reduction). The triple-stakeholder approach successfully generated 1.2 million additional visits to religious sites without compromising user satisfaction. User-centric distance modeling improved both quantity (+9%) and quality (+7%) metrics, while LLM initialization enhanced WMF model performance (+22.6% quantity, +8% quality).

This research contributes to multistakeholder recommender systems by demonstrating that systematic ablation testing can identify multiple improvement directions simultaneously and provide a baseline for future work.

Keywords

Multistakeholder Recommender Systems, Urban Tourism, Sustainability, User Utility, POI Recommendation, Cold Start Problem, Ablation Testing

1 Introduction

In the context of sustainable urban tourism, destination managers face a significant challenge: preventing popular POIs from becoming overcrowded while maintaining high levels of user satisfaction. This creates a conflict between two distinct stakeholders. On the one hand, destination managers seek to reduce overcrowding at popular locations and redistribute visitor flows to peripheral areas. On the other hand, tourists seek personalized recommendations that are biased toward popular sites, often located in city centers.

The core problem addressed in this study is how to mediate these conflicting goals using MRS without compromising the tourists' experiences.[1]

Standard research methods in this field face significant limitations:

- **A/B Online Tests:** Although effective, deploying a system with uncertain effects on the tourism economy is widely considered too risky.
- **Standard Offline Evaluation:** Traditional offline testing using historical data cannot accurately predict user responses to recommended items they did not originally choose.
- **Existing Simulations:** Previous studies have utilized simulations to assess metrics such as diversity and satisfaction. However, these often focus on evaluating exposure (what the MRS recommended) while artificially constraining users to select strictly from recommended lists.

Unlike previous studies that assume users possess full knowledge or are forced to follow recommendations, this study builds upon the framework by Merinov & Ricci [1]. Their work demonstrated that a positive-sum impact is achievable using real-world visitation data (binary visited/non-visited records) from three tourist cities: Rome, Florence, and Istanbul.

We model users' limited awareness of the city and their stochastic decision-making process (based on their prior knowledge and recommended POIs). This allows for the simulation of realistic scenarios where an MRS introduces satisfactory sustainable alternatives to users' awareness sets, enabling choices that may result in higher utility than their original plans.

To mediate between the two stakeholders:

[1]Source code: https://github.com/ofir6bd/UrbanTourismMRS.git

(1) **Destination Utility (v_j):** If a POI is within the top 20% closest to the city center and within the top 20% most popular POIs, the score equals 1, otherwise, 0.
(2) **User Utility (u_{nj}^t):** The user's predicted preference score for a specific POI.

We use $\lambda \in [0, 1]$ as the trade-off parameter in a linear optimization function:

$$R_n(\lambda) = \underset{(x_1,\ldots,x_J):\sum x_i = B}{\arg\max} \sum_j x_j \left((1-\lambda)\, u_{nj}^t + \lambda\, v_j \right) \tag{1}$$

where $\lambda = 0$ means the system acts as a traditional recommender system and promotes user utility only, and conversely $\lambda = 1$ promotes peripheral POIs only without considering user preferences.

The reference paper demonstrated this phenomenon across Rome, Florence, and Istanbul. Furthermore, this study focuses on Rome only.

This study's objectives are to show that improvement can be achieved across multiple directions, evaluated through ablation testing along the following:

User – Preferred Distance from City Center: Instead of treating POI location as a binary attribute (city center vs. periphery), we model crowdedness as a continuous variable, simulating and respecting each user's preferred distance from the city center.

Model – K-Nearest Neighbors (KNN): To better handle sparse data inherent in tourism, where users visit only a tiny fraction of available locations, we implement and evaluate KNN-based models. While the reference study relied on matrix factorization and autoencoder-based models (VAE and WMF), the literature suggests memory-based approaches like KNN are often more effective for binary data (implicit feedback) and cold-start scenarios.

Item – Promoting Third-Party Stakeholder: We explore maximum stakeholder promotion with minimal negative impact on users, introducing a third stakeholder goal. Specifically, we chose to promote religious/spiritual sites beyond just sustainable POIs.

POI System – Reduced Walking and Smart Routing: To increase simulation realism, we incorporate a reduced walking mechanism that simulates each user's current location and recommends POIs based on proximity to the user's current location. Additionally, we add "smart routing" logic that arranges POIs to minimize walking distance between selected POIs, requiring adjustment of user satisfaction evaluation based on physical effort required.

LLM – Enhanced Semantic Initialization: Traditional matrix factorization approaches rely solely on user-item interactions, which are often extremely sparse in tourism contexts. By initializing the WMF model with semantic embeddings derived from POI descriptions, we provide the system with rich contextual knowledge about attractions before collaborative filtering training begins.

2 Related Work

As mentioned in the introduction, in contrast to existing studies that strictly constrain user choices to recommendation lists [6], Merinov & Ricci [1] present a sophisticated method to perform full simulation and evaluation. The key innovation is modeling the user's limited city awareness and expanding their awareness set with recommendations from the MRS, while modeling their choices using a stochastic decision-making process.

Based on a review of the literature, we identified several studies that inspired us to perform our improvements as follows:

2.1 User Improvement

Lim et al. (2015) [2] demonstrate the effectiveness of maximizing user utility by creating personalized tour recommendations based on user visit durations and spatial preferences. Their work shows that time-based user interest, derived from actual visit durations, provides more accurate personalization. We propose a personalized preferred distance from city center derived from actual visit patterns for more accurate personalization.

This approach quantifies spatial preferences by analyzing each user's historical POI visits to determine their preferred distance from the city center, treating POI location as a continuous variable based on distance from the city center. The system penalizes recommendations that significantly deviate from a user's established spatial preference pattern, ensuring recommendations align with their comfort zone (crowded vs. calm areas).

In this case it is distance, but for example in e-commerce it can be preferred price, assuming the users more likely to buy and explore around they preferred price.

2.2 Model Improvement

The simulation models employed by Merinov & Ricci [1] use model-based approaches including matrix factorization and autoencoder models (such as VAE and WMF) to predict user preferences. However, Anwar et al. (2022) [3] demonstrate that memory-based collaborative filtering approaches such as KNNBaseline address cold-start and sparsity issues more efficiently in achieving lower error rates, which are inherent in tourism data where users visit very few locations initially. The KNN model utilizes baseline factors and handles missing ratings more effectively through neighborhood-based predictions.

Additionally, for comprehensive evaluation and reference purposes, we implemented a simple hybrid KNN approach combining both user-based and item-based collaborative filtering.

2.3 Item Improvement

The traditional MRS mediates between users and destination managers. Following our literature review, we note that Burke et al. (2016) [4] demonstrated that recommender systems can serve multiple stakeholders beyond just the traditional user-item-based framework by adding system owner utility simultaneously. Their work established that incorporating additional stakeholder utilities into recommendation algorithms was both feasible and beneficial in real-world scenarios.

In this study, we adapted this multi-stakeholder concept to the tourism domain to evaluate whether MRS recommendations can increase visitation to religious/spiritual sites (such as churches, synagogues, and temples) as a third stakeholder interest, alongside the existing user satisfaction and sustainability goals.

2.4 POI System Improvement

Recent work by Zamiechowska et al. (2025) [5] presents a "routing algorithm" that "dynamically constructs diverse routes featuring curated POI" for real-time mobile deployment.

Although the mechanism differs, this inspired us to implement "Reduced Walking" and "Smart Routing" logics:

Reduced Walking: Starting from the user's current location (center of historical POIs), our MRS penalizes POI recommendations based on distance to user location, biasing recommendations toward closer POIs.

Smart Route: We implement an algorithm that arranges selected POIs to minimize walking distance between consecutive visits. This addresses a limitation in the reference study [1], where POI selection ignored spatial relationships.

Enhanced Evaluation: To evaluate the impact of routing on user satisfaction, we adjust user satisfaction scores based on physical effort by normalizing walking distances and applying the normalized distances to the user utility score, providing more realistic recommendations.

2.5 LLM Initialization Improvement

Chen et al. (2025) [7] demonstrate that "extensive evaluations on real-world datasets validate that LLMInit significantly boosts the performance of various state-of-the-art CF models."

By extracting semantic embeddings from POI textual descriptions (titles, categories, reviews) using pretrained language models to initialize WMF item embeddings, we create superior starting points compared to random initialization. This approach is particularly valuable in tourism contexts, where POI descriptions contain crucial contextual information about location, type, cultural significance, and visitor experience that sparse interaction data alone cannot capture.

3 Method

This chapter presents our methodological contributions that build upon the baseline MRS framework. We describe the motivation and implementation details for each contribution.

3.1 Simulation Process

3.1.1 Data Collection and Preprocessing. The dataset consists of real-world implicit feedback from Flickr (Rome)[2], an application where locals and tourists upload and share their photos, meaning the data is binary (visited/non-visited). To specifically focus on tourism, the data is filtered to distinguish tourists from locals. We identified tourists as users with $\leq$ 14 sessions averaging $>$ 1.25 photos per session.

3.1.2 Initialization. Before the simulation begins, we must estimate the "true" behavior of the users and generate the awareness set (the core innovation of the reference paper).

- **True Preferences (u_n^*):** EASE algorithms are used to estimate the user's actual preferences based on their full history.
- **Awareness Set (A_n^*):** To simulate a realistic "Cold-Start" scenario, the MRS is given only a small subset of data ($t = 2, 3$, or 4 items) regarding the user's history. We expand the user awareness set by adding the k most popular items ($k = 4, 8, 12, 16$, or 20) that users likely know about but did not visit, representing different knowledge levels of the city.

[2]Dataset and original framework: https://github.com/igobrilhante/TripBuilder

- **Behavioral Stiffness (β_n^*):** We calculate a specific β parameter for each user. This defines whether a user is a "maximizer" (strictly follows high utility) or "explorer" (choosing more randomly). This is calibrated by maximizing the log-likelihood of their historical choices.

3.1.3 MRS Recommendations. In this step, the system intervenes in the user's decision process. Based on the limited history (t items) known to the system and the most popular items (k items), the MRS uses VAE and WMF models to predict each POI's score (u_{nj}^t). This score is combined with the POI sustainability score v_j (0 or 1) using the parameter that balances between stakeholders. Using argmax (Equation 1), the system generates eight recommendations (R_n). These expand the user's awareness set.

$$A_{new} = A_n^* \cup R_n(\lambda) \tag{2}$$

3.1.4 Selection. The user chooses an item to visit (Y_n) from this new awareness set. The decision is modeled stochastically using a Multinomial Logit:

$$P_{nj} = \frac{e^{\beta_n u_{nj}^*}}{\sum_{i \in A_{new}} e^{\beta_n u_{ni}^*}} \tag{3}$$

This formula calculates the probability of choosing item j, heavily influenced by the user's specific β_n and the true utility of the item.

3.1.5 Uplift Evaluation. Finally, the method evaluates the success of the intervention using Uplift Metrics compared to a control policy (no recommendations).

- **User Uplift (τ):** Measures the percentage increase in the user's experienced utility.
- **Destination Uplift (η):** Measures the percentage change in site visits. The goal is to identify a λ value where both τ and η are positive (Positive-Sum).

This is explained in depth in Section 4 (Evaluation).

3.2 Ablation Testing Method

To fully understand the system's capabilities and identify directions for improvement, ablation testing is particularly valuable because multistakeholder systems are inherently complex. By systematically testing each improvement in isolation, we can determine which modifications provide genuine value.

To test the maximum capability of the system, since there are many directions to improve it, we perform ablation testing. Each improvement, with its motivation and implementation, is detailed below:

3.2.1 ***User Improvement.***

Motivation: Instead of using POI location as binary (city center vs. periphery) for sustainability purposes only, tourists exhibit varying spatial preferences. Some prefer crowded central areas, while others seek quieter peripheral locations. By modeling each user's preferred distance from the city center as a continuous variable derived from their historical visit patterns, we can provide more personalized recommendations that respect individual spatial comfort zones, potentially improving utility for both stakeholders.

Implementation: We calculate each user's preferred distance by analyzing the average distance of their t items from city center. During recommendation generation, we apply a distance penalty with weight (0.2) to POIs that deviate significantly from this preference. The adjusted user preference:

$$\text{score}_{\text{adj}} = \text{score}_{\text{orig}} - w_d \cdot |d_{\text{poi}} - d_{\text{pref}}| \tag{4}$$

3.2.2 *Model Improvement*.

Motivation: Model selection represents a key direction for MRS improvement. The current models use matrix factorization and autoencoder-based models (WMF & VAE) for preference prediction. However, tourism data exhibits extreme sparsity, as users typically visit only a tiny fraction of POIs, creating challenging cold-start scenarios. Anwar et al. (2022) [3] demonstrated that KNNBaseline achieved the lowest error rates across all cross-validation scenarios, outperforming matrix factorization techniques that struggle with missing ratings in sparse datasets.

Implementation: We implement two KNN-based approaches:

KNN Baseline: (per literature review) A pure neighborhood-based approach using optimal parameters from literature ($k = 21$ neighbors, cosine similarity). The model operates in three stages: First, it calculates baseline predictions using the formula baseline = $\mu + b_u + b_i$, where μ is the global average of the interaction matrix, b_u represents how a user's average rating deviates from the global mean, and b_i captures how an item's average rating deviates from the global mean. Second, it computes cosine similarity between users based on their POI visit patterns to identify the most similar neighbors. Finally, it adds the computed neighbor influence.

HybridKNN: Combines user-based and item-based collaborative filtering with different weighting schemes to find the optimal balance between user and item contribution ($\alpha = 0.25, 0.5, 0.75$) (for reference only):

$$\text{score} = \alpha \cdot \text{score}_{\text{user}} + (1 - \alpha) \cdot \text{score}_{\text{item}} \tag{5}$$

3.2.3 *Item Improvement*.

Motivation: Traditional MRS frameworks mediate between two stakeholders (tourists and destination managers). Extending the framework to a third stakeholder without negatively impacting the existing two stakeholders represents a meaningful improvement for the system's overall output and is potentially applicable to other domains. By introducing religious/spiritual sites as a third stakeholder, we can test whether the MRS can simultaneously satisfy user preferences, destination sustainability goals, and religious site promotion without creating negative trade-offs.

Implementation: We identify religious sites using keyword matching ("church," "temple," "mosque," "basilica," etc.) and apply a penalty with weight (0.2) to non-religious sites during the recommendation process:

$$\text{score} = (1 - \lambda) \cdot u_{nj} + \lambda \cdot v_j - p_{\text{rel}} \tag{6}$$

3.2.4 *POI System Improvement*.

Motivation: In real-world tourism applications, the physical distance between POIs has significant practical implications. To improve the MRS in the direction of POI realism, we must account for physical effort.

Implementation: This includes two phases:

- **Phase 1:** Generating more relevant recommendations based on proximity.
- **Phase 2:** Optimizing the selected POIs to minimize route distance.

Reduced Walking Logic: Since the data does not contain user real-time location, we simulate each user's starting location as the centroid of their known POIs (t items). During the recommendation process, we apply a distance penalty with weight (0.1) to POIs based on their distance from the user's current location (normalized to the range $[0, 1]$):

$$u_{\text{pen}} = u_{nj} - w_d \cdot d_{\text{norm}} \tag{7}$$

Smart Routing Optimization: After users have made their POI selections, we apply a nearest-neighbor routing algorithm. The algorithm iteratively selects the nearest non-visited POI until all chosen locations are arranged in an optimal sequence that minimizes total walking distance between consecutive POI visits.

3.2.5 *LLM-Enhanced Semantic Initialization Improvement*.

Motivation: Traditional matrix factorization models start with random initialization, which is particularly problematic in cold-start scenarios, where achieving meaningful latent representations is challenging. Enriching the latent vectors with LLM-derived semantic knowledge can accelerate model learning by providing meaningful starting points that capture inherent relationships between POIs based on their semantic properties rather than relying solely on interaction patterns.

Implementation: We enhance the WMF model with semantic initialization using LLM embeddings:

POI Description Creation: We generate comprehensive textual descriptions for each POI using an LLM incorporating historical context, architectural features, and cultural significance. For example: "Pantheon – Built under Emperor Hadrian in the 2nd century AD, the Pantheon is a perfectly preserved Roman temple with the world's largest unreinforced concrete dome."

Semantic Embedding Generation: Using the `all-mpnet-base-v2` sentence transformer model, we generate 768-dimensional embeddings that capture semantic relationships between POIs based on their descriptions.

Dimensionality Compression: To match the MRS framework's 32-dimensional latent factors, we apply variance selection to retain the most discriminative semantic directions across all POI embeddings.

Training Process: To preserve the LLM initialization benefits, the WMF model employs a two-phase training approach:

- **Phase 1:** Initialize item factors with compressed LLM embeddings and freeze them for 20 iterations while training user factors.
- **Phase 2:** Fine-tune both user and item factors jointly to optimize collaborative filtering objectives.

4 Evaluation

To systematically isolate and evaluate the contributions of our proposed improvements, we employ a two-phase evaluation methodology:

- **Phase 1:** We measure the percentage change between MRS output with and without recommendations (control policy), utilizing the user and destination uplift metrics detailed in Section 3.1.5.
- **Phase 2:** We execute comparative simulations with and without our proposed improvements.

Our experimental design encompasses 615 simulations with varying parameter combinations: $t \in \{2, 3, 4\}$ representing user history size, $k \in \{4, 8, 12, 16, 20\}$ the number of popular items the user is aware of, and $\lambda \in [0, 1]$ as the trade-off parameter, tested across 41 discrete values.

The objective is to identify λ values where both user-uplift and destination-uplift achieve positive values (Positive-Sum impact). The results, illustrated for example in Appendix Figure 1, demonstrate user uplift versus destination uplift across different λ values. Low λ values favor user preferences while harming destination goals, whereas high λ values promote destination sustainability at the expense of user satisfaction. The optimal operating region lies within the win-win zone (upper-right quadrant), where λ ranges between 0.2 and 0.4.

To quantify performance improvements numerically, we define two evaluation metrics, measured within the win-win zone only:

- **Quantity:** The total number of positive-sum data points.
- **Quality:** The aggregate sum of user uplift and destination uplift values.

For the POI System improvement (Section 3.2.4), we incorporate walking distance into the evaluation framework by normalizing total walking distances and scaling them to the range $[-0.1, +0.1]$, where lower values correspond to higher walking distances, and adding this term to user uplift scores for both the baseline and improved MRS versions. This ensures that reduced walking distances translate into improved user satisfaction metrics.

5 Results

As described in the previous section, results are measured using quantity and quality metrics. The baseline reference, without proposed improvements, achieved 156 points in the win-win quadrant with total aggregates of 23.25.

The results of our proposed improvements are presented in Table 1, showing percentage changes relative to the baseline:

5.1 Results Overview

User Improvement (Appendix Figure 1): Achieved improvements in both quantity and quality metrics.

Model Improvement:

a. KNN Baseline (Appendix Figure 2): Demonstrated improvement in quantity but degradation in quality.
b. Hybrid KNN (Appendix Figure 3): Showed no improvement in either metric.

Item Improvement (Appendix Figure 4): Achieved improvement in quantity and generated 1.2 million additional visits to religious/spiritual POIs without compromising user satisfaction.

POI System Improvement (Appendix Figures 5 and 6): With walking distance incorporated into evaluation metrics:

a. Reduced Walking: Baseline results achieved 191 points in the win-win quadrant, total aggregation of 38, and average walking distance of 16.2 km.
b. Reduced Walking & Smart Routing: Since the evaluation metric incorporates normalized walking distances for both the baseline and the improvement, and because smart routing significantly reduces walking distance, the normalization shifts the baseline to 127 points in the win-win quadrant, with a total aggregation of 15.5 and an average walking distance of 16.2 km.

Both the Reduced Walking and Smart Routing improvements demonstrated enhanced performance in quantity, quality, and walking distance metrics.

LLM Initialization – WMF (Appendix Figure 7): The baseline reference, WMF with random initialization achieved 53 points in the win-win quadrant with total aggregation of 7.574. The WMF with LLM initialization demonstrated improvement in both quantity and quality.

6 Discussion

The results of this study demonstrate that multistakeholder recommender systems can indeed achieve positive-sum outcomes in urban tourism, confirming and extending the foundational work by Merinov & Ricci [1]. Our systematic ablation testing across five proposed improvements reveals varying degrees of success, providing insights into both tourism and broader recommender system applications.

6.1 User (Preferred Distance from City Center)

This improvement achieved gains in both quantity (+9%) and quality (+7%) metrics, demonstrating successful utilization of a hidden user attribute (crowdedness preference) for enhanced personalization. This provides a valuable insight into recommender systems: differentiating users through individual spatial features can significantly improve recommendation quality.

While POI crowdedness is implicitly modeled by WMF and VAE, explicit modeling, especially with controllable impact, proves more effective. This approach generalizes to other domains, for example, in e-commerce, a user's preferred price range can serve as an analogous feature to preferred distance from city center. Users tend to explore and purchase items within their preferred price range, regardless of how much they typically favor such products.

6.2 Model (KNN)

KNN Baseline: yielded mixed results, with improved quantity (+14.7%) but decreased quality (-49%). The global mean anchor, expanded the quantity in the win-win zone but reduced personalization quality.

Hybrid KNN: further confirmed that higher user-based weighting produces better recommendations, reinforcing the importance of user-centric modeling in multistakeholder systems.

6.3 Item (Religious/Spiritual POI Promotion)

The triple-stakeholder framework successfully generated 1.2 million additional visits to religious/spiritual sites without compromising user satisfaction (quality +0.7%, quantity +9%). This demonstrates that additional stakeholder objectives can be incorporated without negative trade-offs, achieving "more gain without pain" outcomes in multistakeholder systems.

6.4 POI System

Incorporating spatial constraints improved results, as the absence of distance considerations was a key limitation in the original system.

Table 1: Results Summary: Comparison of Improvements vs. Baseline

#	Improvement	N Points	Total Aggregation	Walking Distance
	Baseline	156	23.25	N/A
1	User (Distance Preference)	170 (+9%)	24.87 (+7%)	N/A
2	**Model**			
	KNN Baseline	179 (+14.7%)	11.87 (-49%)	N/A
	Hybrid KNN (25% User CF)	8 (-95%)	0.9 (-96%)	N/A
	Hybrid KNN (50% User CF)	24 (-85%)	2.6 (-89%)	N/A
	Hybrid KNN (75% User CF)	27 (-82%)	3 (-87%)	N/A
3	Item (1.2M Visits in Religious/Spiritual POIs)	170 (+9%)	23.4 (+0.7%)	N/A
	POI System Baseline	191	38	16.2 km
4	**POI System**			
	Reduced Walking	214 (+12%)	43.6 (+14.7%)	15.95 km (+1.5%)
	Smart Routing	322 (+253%)	67.6 (+436%)	7.46 km (+54%)
	WMF Baseline (Random Init)	53	7.574	N/A
5	LLM Initialization (WMF)	65 (+22.6%)	8.19 (+8%)	N/A

While the Reduced Walking improvement showed modest distance gains (+1.5%), likely due to cold-start constraints, the smart routing optimization delivered substantial improvements across all metrics. The 54% reduction in walking distance not only enhanced user satisfaction but unexpectedly improved destination uplift, enabling the use of higher λ values that favor sustainability goals.

6.5 LLM Initialization

This approach also leverages a hidden attribute, but unlike the user-centric improvement, it targets semantic relationships between items (POIs). This improvement is specific to the WMF model architecture, enhancing the performance (+22.6% quantity, +8% quality).

Fine-tuning the first phase (frozen iterations) requires careful balance: too few iterations provide insufficient time for user factors to adapt to the semantic item representations, while too many may cause overfitting to the frozen semantic embeddings, diminishing the benefits of collaborative filtering adaptation.

This approach demonstrates how external knowledge sources can enhance traditional collaborative filtering methods, particularly in cold-start scenarios.

7 Conclusion

This research successfully demonstrates that MRS for urban tourism can be systematically improved through ablation testing, extending the foundational positive-sum framework established by Merinov & Ricci (2024) [1]. Through comprehensive evaluation of distinct improvements, this study identifies varying degrees of improvement across all five ablation directions.

Our findings reveal that POI System enhancements deliver the most substantial improvements, with smart routing optimization achieving remarkable gains of +253% in quantity metrics, +436% in quality metrics, and a 54% reduction in walking distances. This demonstrates that incorporating spatial realism into recommender systems not only enhances user experience but amplifies sustainability benefits by enabling higher λ values that favor peripheral POI promotion.

User-centric modeling through preferred distance from the city center consistently improved both quantity (+9%) and quality (+7%) metrics, highlighting the importance of leveraging hidden user attributes for enhanced personalization.

The triple-stakeholder framework proves effective, generating 1.2 million additional visits to religious/spiritual sites without compromising existing stakeholder objectives.

The LLM-enhanced semantic initialization provides a promising direction for incorporating external knowledge through POI semantic embeddings. This approach addresses cold-start challenges in sparse tourism data, achieving +22.6% quantity and +8% quality improvements for the WMF model.

Our model improvements using KNN revealed nuanced results, with KNN Baseline improving quantity (+14.7%) while reducing quality (-49%), suggesting that while neighborhood-based approaches can expand the positive-sum operating region, they may sacrifice personalization quality relative to matrix factorization methods.

This research establishes a foundation and robust baseline for future MRS research, supporting continued innovation in balancing competing stakeholder interests.

8 Limitations

8.1 Data Constraints

Our evaluation relies solely on the Rome Flickr dataset, limiting geographical and cultural generalizability. The decade-old data may not reflect current tourism patterns/behaviors.

8.2 Evaluation Methodology

The offline simulation approach, while safer than live A/B testing, cannot fully account for the offline-online mismatch. User responses to actual recommendations may differ from simulated behaviors.

8.3 Ablation Testing

Our systematic ablation approach tests each improvement in isolation to identify individual contributions. However, this methodology cannot capture potential synergistic effects or conflicts that may emerge when multiple improvements are combined.

9 Future Work

9.1 Dynamic Optimization

Implementing a trade-off parameter (λ) that adapts based on temporal factors such as time of day or season to manage peak tourism hours more effectively.

9.2 Multi-Improvement Integration

Investigating the combined effects of multiple improvements simultaneously, as the current ablation methodology tests improvements in isolation.

9.3 Crowding Feedback Loops

Developing crowding simulation by calculating POI visit times and generating dynamic crowding data. This data would feed back into the recommendation algorithm to prevent overcrowding proactively.

9.4 Cross-Domain Application

Extending the limited awareness modeling and positive-sum simulation framework to domains beyond tourism, such as e-commerce.

Acknowledgments

We thank the authors of the original study by Merinov & Ricci (2024) [1], for their foundational work on MRS for urban tourism.

References

[1] Pavel Merinov and Francesco Ricci. 2024. Positive-Sum Impact of Multistakeholder Recommender Systems for Urban Tourism Promotion and User Utility. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24)*, October 14–18, 2024, Bari, Italy. ACM, New York, NY, USA.

[2] Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2015. Personalized tour recommendation based on user interests and points of interest visit durations. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI '15)*, Vol. 15, 1778–1784.

[3] Tehmina Anwar, V. Uma, Md Imran Hussain, and Mohana Pantula. 2022. Collaborative filtering and kNN based recommendation to overcome cold start and sparsity issues: A comparative analysis. *Multimedia Tools and Applications* 81 (2022), 35693–35711.

[4] Robin D. Burke, Himan Abdollahpouri, Bamshad Mobasher, and Trinadh Gupta. 2016. Towards multi-stakeholder utility evaluation of recommender systems. In *UMAP (Extended Proceedings)*, 750.

[5] Julita Zamiechowska, Julia Neidhardt, and Wolfgang Wörndl. 2025. CiRi-Engine: POI Recommender System for Diverse and Balanced Walking Tours. In *Workshop on Recommenders in Tourism (RecTour 2025)*, co-located with the 19th ACM Conference on Recommender Systems, Prague, Czech Republic.

[6] Nada Ghanem, Stephan Leitner, and Dietmar Jannach. 2022. Balancing consumer and business value of recommender systems: A simulation-based analysis. *Electronic Commerce Research and Applications* 55 (2022), 101195. https://doi.org/10.1016/j.elerap.2022.101195

[7] Junhan Chen, Xiaoliang Wang, Yaojun Li, Mingming Zhang, and Hao Liu. 2025. LLMInit: A Free Lunch from Large Language Models for Selective Initialization of Recommendation. *arXiv preprint arXiv:2503.01814*.

Appendix

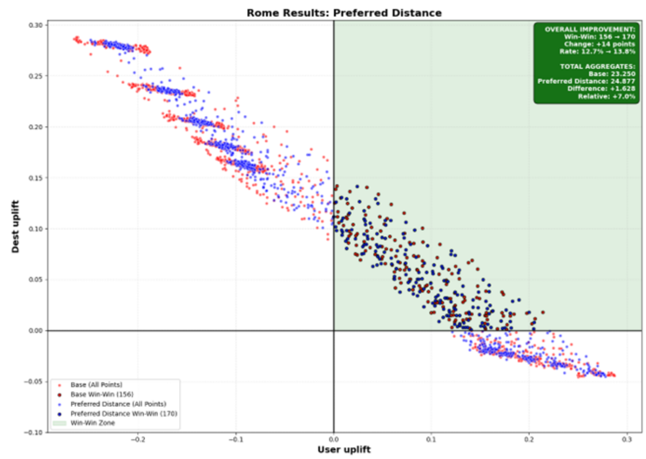

Figure 1: Improvement 1 – User Preferred Distance from City Center.

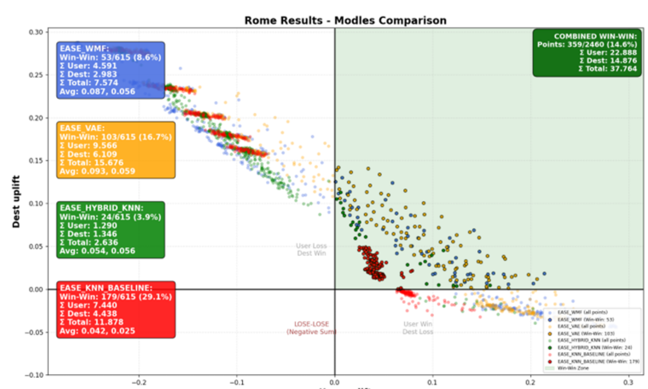

Figure 2: Improvement 2 – KNN Baseline and Hybrid KNN (50%) vs. Baseline Results (WMF & VAE).

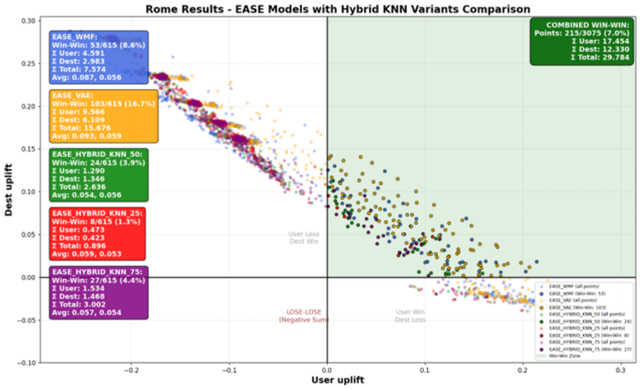

Figure 3: Hybrid KNN (25%, 50%, 75%) vs. Baseline Results (WMF & VAE).

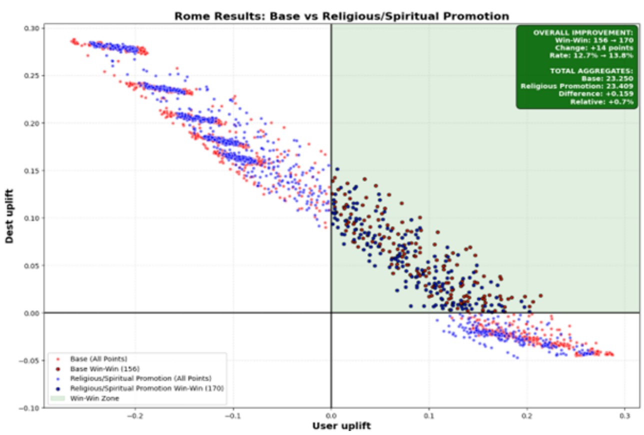

Figure 4: Improvement 3 – Promote Religious/Spiritual POIs.

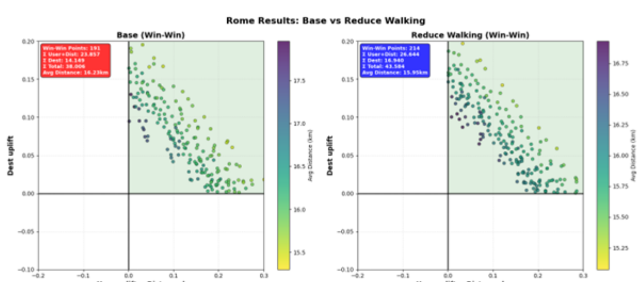

Figure 5: Improvement 4 – Reduced Walking vs. Baseline Results.

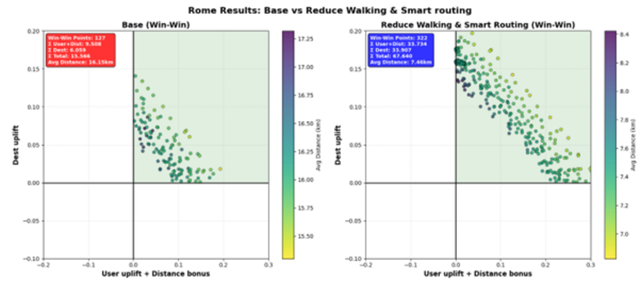

Figure 6: Improvement 4 – Reduced Walking & Smart Routing vs. Baseline Results.

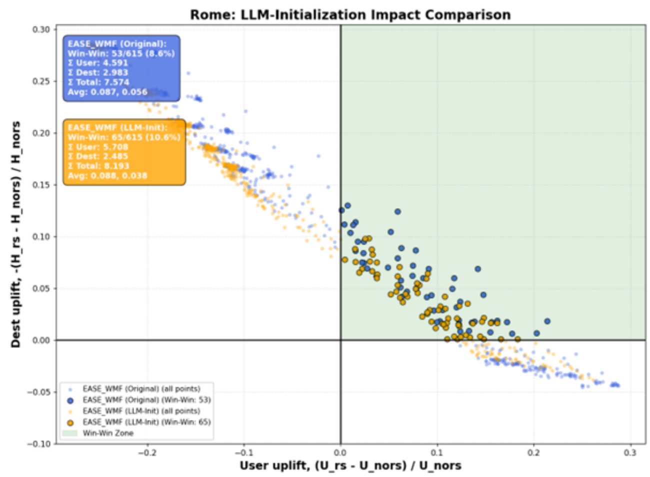

Figure 7: Improvement 5 – LLM WMF Initialization.