Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9100

MultiModal Adaptive Fusion for Itinerary Generation Robust to Missing Modalities

Minmin Zhang
Tianjin University of Technology and Education, Tianjin, China
Email: 2871085125@qq.com

How to cite this paper: Zhang, M. (2026). MultiModal Adaptive Fusion for Itinerary Generation Robust to Missing Modalities. Journal of Computer Science and Frontier Technologies, 4(2), 20–30. ISSN Print: 3104-4204, ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9100
Published: 2026-09-28

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Abstract

Tour route recommendation aims to generate feasible travel sequences based on user needs and time budgets. Most existing tour route generation models rely heavily on trajectory sequence information and underutilize multimodal data of attractions, leading to limited adaptability to real-world scenarios with missing modalities and a lack of explicit modeling of tourists' diverse interests and visiting order preferences. To address these issues, this paper proposes a multimodal adaptive fusion method for tour route generation that is robust to missing modalities. It models user preferences through a gated fusion mechanism, multi-interest decoupling, and a recent-visit memory gating module, and generates feasible routes by incorporating pairwise order loss and beam search with duration penalty. Comparative experiments and ablation studies conducted on real-world tourism datasets verify the effectiveness of the proposed model.

Keywords

Tour route recommendation; Sequence generation; Multi-modal fusion; Multi-interest decoupling; Beam search

1. Introduction

With the rapid development of the mobile Internet and the increasing popularity of Global Positioning System (GPS) devices, massive digital footprints left by users on various platforms provide unprecedented opportunities for in-depth understanding of human mobility patterns and supporting downstream applications such as itinerary planning and location recommendation. For this reason, tourism recommendations have continuously captured the attention of both the industry and academia over the past decade [1].

Tourism recommendation aims to plan an orderly itinerary for users based on their available travel time, designated starting and ending points, and the number of attractions they wish to visit. On one hand, the interests and budget constraints of different tourists often vary greatly; on the other hand, factors such as spatial

distance, the attractiveness of the attractions themselves, the duration of the visit, the overall diversity of the route [2], and even the waiting time for popular attractions [1] are all within the scope of consideration for trip planning.

In response to these challenges, existing studies fall into two main categories. The first category formulates itinerary planning as the Orienteering Problem (OP) and solves it to maximize the total revenue of routes [3]. Nevertheless, OP is NP-hard. The computational cost of obtaining exact solutions grows exponentially with the number of candidate POIs, which makes it difficult to scale up to larger-scale datasets [4]. Moreover, these methods cannot explicitly capture users' interests in attractions and visiting patterns, resulting in relatively low personalization and failing to fully meet the demands of individual tourists.The second approach learns tourist preferences from data. Early learning methods (such as those based on Markov chains [5]) combine location, POI categories and historical trajectories, and utilize transition probabilities to model the sequential dependencies between attractions. In recent years, deep neural networks have been widely adopted. For instance, TRED [6] uses recurrent neural networks (RNN) combined with attention mechanisms to model the temporal and sequential patterns of user trips. There are also works that utilize POI embeddings to model user travel preferences [7]. POIBERT [8], DeepTrip [9], and CTLTR [10] respectively utilize Transformer, adversarial learning, and self-supervised contrastive learning to uncover the internal transfer patterns within trajectories. BTREC [11] is based on the BERT architecture and uses a masked language model to iteratively complete the intermediate POIs to achieve trip generation under the constraints of starting and ending points; there are also works that first learn context-aware POI embeddings and then generate routes through planning [12].

Although deep-learning-based methods eliminate manual feature engineering and can effectively capture complex human mobility patterns, most of them still rely solely on historical trajectories. When trajectory data is sparse, it is difficult to adequately mine tourist preferences. Multimodal information such as textual and visual content of attractions can provide complementary information to mitigate the drawbacks caused by sparse trajectories. Nevertheless, existing studies still exhibit obvious limitations in utilizing multimodal information [13].

Tourism route generation is a typical sequence-generation task [14]. To address the above-mentioned drawbacks, this paper proposes a multimodal adaptive fusion for itinerary generation robust to missing modalities.

The main contributions of this article are summarized as follows:

1. To address the issue of incomplete text and image modalities in real tourism data, an adaptive multimodal fusion mechanism based on modal masking is designed, enabling the model to robustly utilize available information even when modalities are missing.
2. Multi-interest decoupling and recent visit memory gating are proposed to

balance the diversified long-term interests of tourists and their recent decision-making reliance, enhancing the personalization of recommendations.

3. Trip planning is modeled as an auto-regressive sequence generation task, and a beam search with duration constraints is used to ensure the feasibility and quality of the generated routes. Experimental verification on real tourism datasets demonstrates the effectiveness of the proposed method.

2. Preliminary Knowledge

2.1. Representation of Attractions

Let $L = \{l_1, l_2, \dots, l_{|L|}\}$ represent the set of tourist attractions in the target city, with a total of $|L|$ attractions. Each attraction $l \in L$ is associated with three types of information.

Structured Information: This includes the scenic spot identifier, the city it belongs to, the main category, the latitude and longitude coordinates (lat_l, lon_l), and the suggested visiting duration $visit(l)$, which together constitute the structured attribute vector s_l of the scenic spot.

Visual feature: Visual features are extracted from scenic spot images by a visual encoder, resulting in a visual feature vector v_l.

Text feature: The text feature vector t_l is obtained by extracting text features from the description text of the scenic spot through a text encoder.

2.2. Route Definition

A tourist route is defined as an ordered sequence of attractions:

$$r = (l_1, l_2, \dots, l_N)$$

Here, l_1 represents the starting point, l_N represents the destination point, and N represents the total number of scenic spots included in the route. Moreover, the route does not contain any duplicate scenic spots, that is $l_i \neq l_j, \forall i \neq j$.

Given route r, the total duration is composed of the time spent visiting each attraction and the time taken for transportation between attractions:

$$Time(r) = \sum_{l \in r} v\,isit(l) + \sum_{i=1}^{N-1} \left(\frac{60 \cdot \rho \cdot d(l_i, l_{i+1})}{speed} + overhead \right)$$

Among them, $visit(l)$ represents the duration of visiting the scenic spot l, $d(\cdot,\cdot)$ is the spherical Haversine distance based on latitude and longitude, $speed$ is the set driving speed, ρ is the road condition coefficient, and $overhead$ is the fixed cost for each transfer segment.

2.3. Formalization of the Problem

Given the user query $Q = (l_s, l_e, N)$, where l_s represents the starting point, l_e represents the destination point, N represents the number of target visited points,

and T represents the time budget, the model needs to generate a feasible route:

$$r = (l_s, l_2, \ldots, l_{N-1}, l_e)$$

This paper defines $p_t = (l_s, l_2, \ldots, l_t)$ as the access prefix of the route at step t. Given the access prefix p_t, query condition Q, and model parameters θ, the conditional probability of selecting the next attraction l_{t+1} is:

$$l_{t+1} = \operatorname{argmax}_{l \in L} P(l \mid p_t, Q, \theta)$$

A complete route is constructed through stepwise auto-regressive prediction, and in the decoding stage, candidate attractions are constrained and filtered according to the time budget.

3. Methods

3.1. Multimodal Adaptive Fusion

For a scenic spot l_i, its structured representation S_i is obtained by summing three components: the learnable scenic spot embedding e_i, category information, and geographical information:

$$S_i = e_i + W_c \cdot Emb_c(c_i) + W_g \cdot g_i$$

Here, c_i represents the category number of the scenic spot, $Emb_c(\cdot)$ represents the category embedding, g_i represents the standardized longitude and latitude coordinates, W_c and W_g are the projection matrices for category and geography, respectively.

Given a query Q, the starting point representation t_{l_s} and the ending point representation t_{l_e} are obtained from the structured representation table. They are concatenated with the target length embedding $Emb_{len}(N)$ and then linearly transformed and passed through the tanh activation to obtain the query vector:

$$q = \tanh\left(W_q \cdot [t_{l_s} \oplus t_{l_e} \oplus Emb_{len}(N)]\right)$$

Here, $\oplus$ represents vector concatenation. The query vector aggregates the departure point, destination and expected scale of this trip.

Given the structured representation table S, this paper further integrates visual features and text features. Let the visual feature matrix be F_v and the text feature matrix be F_t, which are respectively linearly projected to the hidden dimension:

$$F_v' = W_v F_v, \quad F_t' = W_x F_t$$

The modal mask m records the availability of each modality for each scenic spot. When a certain modality is missing or discarded by modal dropout, the corresponding channel is set to zero. The structured representation, visual features, and text features are regarded as three parallel channels, and they are adaptively

fused through a learnable gated network.

Let the k-th component of the gated input be $\alpha_k = \left[W_{gate}(S \oplus F_v' \oplus F_t' \oplus m)\right]_k$, where W_{gate} is the weight. Then the normalized gated weights and the fused representation are:

$$g_k = \frac{\exp(\alpha_k)}{\sum_{k'=1}^{3} \exp(\alpha_{k'})}, \quad S_f = \sum_{k=1}^{3} g_k S^{(k)}$$

Here, $S^{(1)} = S$, $S^{(2)} = F_v'$, $S^{(3)} = F_t'$ represent the representations of the structured, visual and textual channels respectively. g_k is the normalized gated weight for the corresponding channel, and layer normalization can be optionally applied at the fusion output.

To depict the diverse tourism interests of users, this paper designs a multi-interest decoupling head. Each route is divided into several prefix samples. For route r, at position t ($1 \le t < N$), a sample is constructed with prefix $p_t = (l_s, \ldots, l_t)$, prediction target l_{t+1}, future targets $l_{t+1}, \ldots, l_e$, starting point l_s, ending point l_e, and target length N.

For the current visited prefix p_t, the fused representation of each position POI is fed into a layer of GRU encoder in access order to obtain the hidden state sequence $h_1, \ldots, h_t$ of the prefix. On this basis, assuming the number of interest heads is K, a set of learnable interest query vectors $\{u_k\}_{k=1}^{K}$ is used to represent different interest prototypes. For the k-th interest head, the attention weights of each position of the prefix are calculated:

$$\alpha_{k,j} = \frac{\exp\left(\langle h_j, u_k \rangle\right)}{\sum_{j'=1}^{t} \exp\left(\langle h_{j'}, u_k \rangle\right)}$$

Then, the k-th interest representation $z_k = \sum_{j=1}^{t} \alpha_{k,j} h_j$ is obtained. Subsequently, based on the query vector q, each interest representation is scored and normalized to obtain the interest weight:

$$w_k = \frac{\exp(\langle z_k, q \rangle)}{\sum_{k'=1}^{K} \exp(\langle z_{k'}, q \rangle)}, \quad z = \sum_{k=1}^{K} w_k z_k$$

The model can dynamically amplify relevant interests and suppress irrelevant ones based on the starting point, destination and scale of the current trip. When the adaptive interest mode is enabled, the model only activates the first $\min(t, K)$ interest heads, so that the number of active interests matches the number of visited steps.

To constrain multiple interests from converging, this paper introduces two auxiliary losses. The diversity loss is based on the squared cosine similarity between two interest representations, penalizing the tendency of different interests to become consistent; the balance loss penalizes the deviation of the interest weight

distribution from the uniform distribution, encouraging the model to fully utilize multiple interest heads.

$$L_{div} = \frac{1}{K(K-1)} \sum_{k \neq k'} \left(\frac{\langle z_k, z_{k'} \rangle}{\| z_k \| \| z_{k'} \|} \right)^2$$

$$L_{bal} = \frac{1}{K} \sum_{k=1}^{K} \left(w_k - \frac{1}{K} \right)^2$$

3.2. Memory Gate Control

In addition to multiple interests, tourists' next decisions are also strongly influenced by their recently visited attractions. This paper introduces a recent-visit memory gate control mechanism. Taking the last-step hidden state h_t, the interest aggregation representation z, and the query vector q as inputs, after a layer of linear transformation and sigmoid activation, the gate control value is obtained:

$$\gamma = \sigma(W_r[h_t \oplus z \oplus q]), \quad z' = \gamma h_t + (1-\gamma)z$$

Here, γ controls the extent to which the model leans towards recently visited attractions. Finally, the gated representation is concatenated with the query vector and projected to obtain the context:

$$c = \tanh(W_h[z' \oplus q])$$

This model uses position-based cross-entropy as the main supervisory signal to predict the next visited attraction:

$$L_{ce} = -\frac{1}{|B|} \sum_{(p_t, l_{t+1}) \in B} \log P(l_{t+1} \mid p_t, q)$$

Here, B represents the batch-level prefix sample set, and $P(\cdot)$ is obtained by applying softmax to the scoring vector.

Besides cross-entropy, the training objective also includes multi-interest diversity loss, balance loss, and pairwise order loss:

$$L = L_{ce} + \lambda_{div} L_{div} + \lambda_{bal} L_{bal} + \lambda_{order} L_{order}$$

where λ_{div}, λ_{bal}, λ_{order} are weight hyperparameters.

The pairwise order loss explicitly learns the preference for the sequence of visits between attractions. Given the prefix p_t and the target l_{t+1}, this paper uses two sets of learnable projections to characterize the preceding attractions and the subsequent attractions respectively, and constrains the order score of the preceding and target pair to be higher than the reverse score:

$$L_{order} = E_{(p_t, l_{t+1})} \left[\frac{1}{t} \sum_{j=1}^{t} \text{softplus} \left(-\langle W_L s_j, W_R s_{l_{t+1}} \rangle - \langle W_R s_j, W_L s_{l_{t+1}} \rangle \right) \right]$$

Here, s_j represents the fused expression at position j in the prefix, while W_L and

W_R represent the left and right sequential projections respectively. The average is calculated over the valid positions within the prefix.

3.3. Route Generation

To strike a balance between generation quality and time constraints, this paper implements beam search with duration constraints. Let T be the time budget set by the user. During the route generation process, several partial routes are maintained; at each step, the top k candidate attractions with the highest scores are expanded for each partial route, and the complete duration and timeout penalties of the expanded candidate routes are calculated:

$$violation = \max\left(0, \frac{Time(r)}{T} - 1\right)$$

The score of each candidate path consists of three components: normalized logarithmic probability, optional pairwise order bonus, and timeout penalty. During the intermediate step expansion phase, the logarithmic probability is normalized by subtracting one from the length of the candidate path:

$$score_{expand} = \frac{\log P(route)}{|route| - 1} + \lambda_{pw} \cdot order(r) - \lambda_c \cdot violation$$

In the final convergence stage after end point correction, the logarithmic probability is normalized by subtracting two from the length of the candidate route:

$$score_{final} = \frac{\log P(route)}{|route| - 2} + \lambda_{pw} \cdot order(r) - \lambda_c \cdot violation$$

Here, λ_{pw} represents the pairwise order weight, and λ_c represents the constraint penalty weight.

$order(r) = \frac{1}{\binom{N}{2}} \sum_{1 \le i < j \le N} score(l_i, l_j)$ uses order projection to calculate the average order score for all ordered pairs of scenic spots within the entire route. The two stages use different normalization denominators. Essentially, both subtract the two definite nodes (the starting point and the ending point), and only normalize the intermediate prediction steps; since routes in the expansion stage do not include the ending point while routes in the convergence stage incorporate the ending point, their lengths are defined differently.

Given the context vector c and the scene representation table S_f, the model's score for each candidate scene is given by the inner product plus the output bias:

$$s_j = c^{\top} S_f[j] + b_j$$

Here, b_j represents the output bias of scenic spot j.

During the generation process, this paper limits candidate scenic spots to the set of spots within the same city as the starting point through city candidate masks, in order to reduce the search space and ensure the geographical feasibility of the route. After the beam search is completed, the feasible route that meets the time budget and has the highest score is returned first.

4. Experiment

4.1. Dataset

Experiments are conducted on two real-world multimodal tourism datasets: the Beijing-Tianjin dataset from China and the Melbourne dataset from Australia. The Beijing-Tianjin subset covers 268 real scenic spots. Each attraction is associated with three types of information: structured attributes , textual descriptions, and visual images. All attractions are uniformly mapped to a shared set of POI codes. For route supervision data, only attractions that can be mapped to this vocabulary are retained, and each route is required to contain at least two valid scenic spots (no fewer than three for formal-use data). A modality mask is provided for each POI to indicate the availability of text and visual modalities, simulating real-world conditions where partial modal information may be missing.

4.2. Evaluation Indicators

This paper employs two core indicators commonly used in tourism route recommendation.

Route-F1: Ignores the order of visits and treats the route as a set of POIs. It calculates the F1 value of the set difference between the predicted route and the reference route, focusing on measuring the matching degree of the set of attractions.

Pairs-F1: Divides the route into a set of ordered POI pairs that satisfy $i < j$. It calculates the F1 value of the ordered pair set of the predicted route and the reference route. This indicator takes into account both the selection of POIs and the order of visits, and is a more rigorous metric for evaluating route generation quality.

4.3. Baseline Models

CTLTR [10]: A tourism recommendation model based on contrastive trajectory learning. It generates subsequences by recursively splitting the trajectory for data augmentation and designs a contrastive loss based on mutual information maximization to learn POI transition patterns and user interests.

BTRec [11]: A personalized tourism trajectory recommendation model based on BERT. Given only a pair of source and destination POIs, it encodes the user's historical trajectory and demographic information into BERT and uses a masked language model to iteratively complete the intermediate POIs to generate personalized trips.

DeepTrip [9]: An end-to-end route recommendation model based on the RNN autoencoder backbone. It learns the distribution of serialized POIs through an auxiliary neural network in an adversarial manner, modeling the similarity between queries and trajectories.

Ours: A route generation model with multimodal adaptive fusion. It uses gated fusion, multi-interest decoupling, and beam search with duration constraints to handle modal missingness and generate trip routes that satisfy multiple constraints.

4.4. Experimental Setup

The hidden dimension of the model is set to 64, the category embedding dimension is 16, and the number of interest heads is set to 2. The maximum route length is limited to 20. Dropout is set to 0.20 and modality dropout is 0.10. The beam width is 4. The batch size is 32, and the maximum training epoch is 100. Early-stopping is adopted with a patience of 20 epochs. The optimizer is AdamW with a learning rate of 0.001 and weight-decay of 0.001, and gradient clipping value is 5.0. Under outer cross-validation, model performance is evaluated on the held-out validation fold after each epoch. Training terminates if no improvement is observed within 20 consecutive epochs.

4.5. Experimental Results

Performance comparison on Beijing-Tianjin and Melbourne datasets

Model	Metric	Datasets	
3-4		Beijing-Tianjin	Melbourne
CTLTR	Route-F1	0.6495 ± 0.1118	0.5781 ± 0.0058
	Pairs-F1	0.3586 ± 0.1627	0.2688 ± 0.0090
DeepTrip	Route-F1	0.6615 ± 0.1180	0.5785 ± 0.0063
	Pairs-F1	0.3778 ± 0.1741	0.2684 ± 0.0110
BTRec	Route-F1	0.6653 ± 0.1172	0.5793 ± 0.0032
	Pairs-F1	0.3816 ± 0.1762	0.2705 ± 0.0057
Ours	Route-F1	$\mathbf{0.6684 \pm 0.1130}$	$\mathbf{0.5829 \pm 0.0029}$
	Pairs-F1	$\mathbf{0.3818 \pm 0.1621}$	$\mathbf{0.2722 \pm 0.0051}$

To evaluate the route generation capability of the proposed model, comparative experiments are conducted on real tourism datasets with three baseline models: CTLTR, DeepTrip, and BTRec. The experimental results are shown in Table 1.

CTLTR relies on contrastive trajectory learning to discover transfer patterns between POIs, but lacks the utilization of multimodal information, resulting in relatively lower overall indicators. DeepTrip models the trajectory distribution using an adversarial autoencoder, and BTRec uses the BERT masking mechanism to complete routes under the constraints of starting and ending points. Compared with CTLTR, both DeepTrip and BTRec achieve certain improvements. The proposed method introduces multi-modal adaptive fusion, multi-interest decoupling modeling, and paired sequential loss, achieving improvements across all evaluation indicators.

4.6. Ablation Experiments

To verify the contribution of each core component to model performance, this paper conducts component-wise ablation experiments, and the experimental results are shown in Table 2.

Results of ablation experiments on the Beijing-Tianjin dataset

Variant	Route-F1	Pairs-F1
w/o Memory Gate	0.6606 ± 0.1142	0.3726 ± 0.1647
w/o Multimodal	0.6638 ± 0.1173	0.3780 ± 0.1735

w/o Order	0.6665 ± 0.1137	0.3796 ± 0.1634
Ours	$\mathbf{0.6684 \pm 0.1130}$	$\mathbf{0.3818 \pm 0.1621}$

Compared with the full model, removing the recent-visit memory gating module results in obvious degradation of all metrics, which confirms the vital role of this mechanism in balancing the influences of long-term multi-interest representations and short-term visiting behaviors on scenic-spot selection. When the multimodal fusion module is removed, the model relies solely on structured data, and substantial performance loss can be observed. This demonstrates that visual and textual modalities provide complementary features for modeling user preferences. In contrast, removing pairwise-order loss deprives the model of explicit order supervision and leads to moderate performance decline, which reflects the auxiliary benefit of explicit order supervision for learning reasonable tourist visiting sequences. Equipped with all the above components, the full model achieves optimal performance, which verifies the practical effectiveness of each core module.

5. Conclusion

This paper conducts research on the tourist route generation task, proposing a multi-modal adaptive fusion method for tourist route recommendation. Through multi-interest decoupling heads to depict the diverse tourism preferences of tourists, combined with diversity loss and balance loss to avoid the convergence of interest representations; at the same time, a recent visit memory gating mechanism is designed to balance the influence of long-term interests and recent visit records on the selection of the next scenic spot. In the training process, paired sequential loss is introduced to explicitly learn the access patterns between scenic spots; in the decoding stage, beam search with duration penalty is used to generate a complete tourist route that meets the time budget constraint. Adequate experiments are conducted on the real dataset. Future work will further study more robust multi-modal pre-training of scenic spot representations, improve the decoding strategy to reduce error propagation, and introduce user profile information to further enhance the personalized ability of route recommendations.

References

[1] Halder, S., Lim, K.H., Chan, J., et al. (2022) POI Recommendation with Queuing Time and User Interest Awareness. Data Mining and Knowledge Discovery, 131.

[2] Xu, T., Trajcevski, G., Kim, J.S., et al. (2020) Semantically Diverse Path Search. Proceedings of the 21st IEEE International Conference on Mobile Data Management (MDM), Versailles, IEEE, 69–78.

[3] Friggstad, Z., Gollapudi, S., Kollias, K., et al. (2018) Orienteering Algorithms for Generating Travel Itineraries. Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM), New York, ACM, 180–188.

[4] Sohrabi, S., Ziarati, K., Keshtkaran, M. (2020) A Greedy Randomized Adaptive Search Procedure for the Orienteering Problem with Hotel Selection. European Journal of Operational Research, 283(2), 426–440.

[5] Cheng, C., Yang, H., Lyu, M.R., et al. (2013) Where You Like to Go Next: Successive

Point-of-Interest Recommendation. Proceedings of the 23rd International Joint Conference on Artificial Intelligence, AAAI, 2605–2611.

[6] Zhou, F., Wu, H.T., Trajcevski, G., et al. (2020) Semi-Supervised Trajectory Understanding with POI Attention for End-to-End Trip Recommendation. ACM Transactions on Spatial Algorithms and Systems, 6(2), 13:1–13:25.

[7] Ho, N.L., Lim, K.H. (2021) User Preferential Tour Recommendation Based on POI-Embedding Methods. Proceedings of the International Conference on Web Information Systems Engineering, 46–48.

[8] Ho, N.L., Lim, K.H. (2022) POIBERT: A Transformer-Based Model for the Tour Recommendation Problem. 2022 IEEE International Conference on Big Data, IEEE, 5925–5933.

[9] Gao, Q., Zhou, F., Zhang, K., et al. (2019) DeepTrip: Adversarially Understanding Human Mobility for Trip Recommendation. Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 110.

[10] Zhou, F., Wang, P., Xu, X., et al. (2022) Contrastive Trajectory Learning for Tour Recommendation. ACM Transactions on Intelligent Systems and Technology, 13(4), 125.

[11] Ho, N.L., Lim, K.H., Lee, R.K.W. (2023) BTRec: BERT-Based Trajectory Recommendation for Personalized Tours. Proceedings of the 1st Workshop on Recommender Systems for Tourism (RecTour@RecSys), 18.

[12] He, J., Qi, S., Ramamohanarao, K. (2019) A Joint Context-Aware Embedding for Trip Recommendations. IEEE 35th International Conference on Data Engineering (ICDE), IEEE, 112–123.

[13] Li, M.L., Lim, K.H., Guo, T., et al. (2023) A Transformer-Based Framework for POI-Level Social Post Geolocation. Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 588–604.

[14] Chen, X., Reibman, A., Arora, S. (2022) Sequential Recommendation Model for Next Purchase Prediction. Technical Report.