A Appendix

A.1 Notations

Key notations used in the paper are summarized in Table 6.

Table 6: Main notations and their definitions.

Notation	Definition
Tr_{in}	the input trajectory
Tr_{au}	the augmented long-term trajectory
p_i, c_i, t_i	POI, POI category and time of a check-in
$\mathbf{X}_{in}^{p}$	embeddings of POI sequence for Tr_{in}
$\mathbf{X}_{in}^{c}$	embeddings of category sequence for Tr_{in}
$\mathbf{X}_{au}^{p}$	embeddings of POI sequence for Tr_{au}
$\mathbf{X}_{au}^{c}$	embeddings of category sequence for Tr_{au}
f_θ	RNN trajectory encoder
f_ϕ	temporal-aware transformer trajectory encoder
z_{in}^{θ}	representation of Tr_{in} obtained by f_θ
z_{au}^{ϕ}	representation of Tr_{au} obtained by f_ϕ
z_{in}^{ϕ}	representation of Tr_{in} obtained by f_ϕ
z_{au}^{θ}	representation of Tr_{au} obtained by f_θ

A.2 Formulations and Architecture

RNN Trajectory Encoder

The architecture of LSTM unit consists of a memory cell and three gates with the flow of information:

$$\begin{aligned} i_t &= \sigma\left(\mathbf{W}_i x_t^* + \mathbf{U}_i h_{t-1} + \mathbf{V}_i c_{t-1} + b_i\right), \\ o_t &= \sigma\left(\mathbf{W}_o x_t^* + \mathbf{U}_o h_{t-1} + \mathbf{V}_o c_{t-1} + b_o\right), \\ f_t &= \sigma\left(\mathbf{W}_f x_t^* + \mathbf{U}_f h_{t-1} + \mathbf{V}_f c_{t-1} + b_f\right), \end{aligned} \tag{10}$$

where i_t, f_t and o_t are the input gate, forget gate and output gate respectively, $\mathbf{W}_*$ and b_* are the learnable parameters and bias vectors, $\sigma()$ is the sigmoid activation function, x_t^* represents the embedding of a check-in in POI sequence or category sequence in Eq. (2).

The memory cell c_t is updated by the previous memory unit c_{t-1} with the current input as:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(\mathbf{W}_c x_t^* + \mathbf{U}_c h_{t-1} + b_c\right), \tag{11}$$

where $\odot$ is the element-wise product.

We take the hidden embedding vector h_t at the last time as the sequence representation, and its update rules are as follows:

$$h_*^{\theta} = h_t = o_t \odot \tanh\left(c_t\right). \tag{12}$$

Trajectory representation z^{θ} is obtained by:

$$z^{\theta} = MLP(\alpha \cdot h_p^{\theta} + (1-\alpha) \cdot h_c^{\theta}). \tag{13}$$

Temporal-aware Transformer Trajectory Encoder

The original transformer is an encoder-decoder architecture (Vaswani *et al.*, 2017). We select its encoder part and revise it as our trajectory encoder. Specifically, it mainly includes two components: token embedding layer and stacked self-attention layer. First, the input POI sequence or category sequence are embedded as:

$$h_0 = [x_1^*\mathbf{E}; x_2^*\mathbf{E}; \cdots; x_m^*\mathbf{E}] + \mathbf{PE} \tag{14}$$

where $\mathbf{E}$ are learnable parameters, x_i^* represents the embedding of the i-th check-in in POI sequence or category sequence in Eq. (2), $\mathbf{PE}$ is our temporal-aware position encoder designed in Eq. (4).

The stacked self-attention layer consists of alternating layers of multi-head self-attention (MSA) and MLP blocks:

$$\begin{aligned} h_l' &= \mathrm{MSA}\left(\mathrm{LN}\left(h_{l-1}\right)\right) + h_{l-1}, \quad & l = 1, 2, \ldots, L \\ h_l &= \mathrm{MLP}\left(\mathrm{LN}\left(h_l'\right)\right) + h_l', & l = 1, 2, \ldots, L \end{aligned} \tag{15}$$

where LN is Layernorm operation which be applied before every block, and residual connection follows every block. L is the number of layers.

To learn the representation for a sequence, we employ max pooling to process the embeddings of last layer tokens:

$$h_*^{\phi} = \text{Max-Pooling}(h_1^L, h_2^L, \cdots, h_m^L). \tag{16}$$

Trajectory representation z^{ϕ} is also obtained by:

$$z^{\phi} = MLP(\beta \cdot h_p^{\phi} + (1-\beta) \cdot h_c^{\phi}). \tag{17}$$

A.3 Experimental Settings

Detailed hyperparameters settings are shown in Table 7. All experiments are conducted on a machine with Intel Xeon Gold 6126 @2.60GHz 12 cores CPU and 8 × NVIDIA Tesla V100-SXM2 (16GB Memory) GPU.

Table 7: Hyperparameters setting.

Hyperparameters	f_θ	f_ϕ
POI embedding dimension	512	512
category embedding dimension	512	512
time embedding dimension	128	128
check-in embedding dimension	512	512
Hidden size	1024	512
Layers	1	3
FFN inner hidden size	\	1024
Attention heads	\	8
Attention head size	\	64
Training epochs	30	
Batch size	512	
Adam ϵ	1e-8	
Adam β	(0.9,0.999)	
Initial learning rate	1e-3	
Learning rate schedule	multiplicative (5,0.9)	
Dropout	0.1	
λ	10	
T	10	
k	8	

A.4 More Detailed Experimental Results

The results of MainTUL with respect to temperature T and hyperparameter λ on Weeplaces are shown in Figure 5.

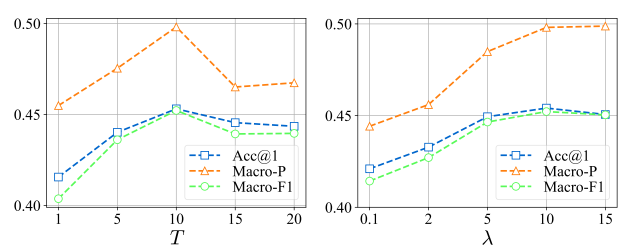

Figure 5: Parameter sensitivity *w.r.t.* T and λ on Weeplaces.