Next Point-of-Interest Recommendation with Auto-Correlation Enhanced Multi-Modal Transformer Network

Yanjun Qin
Beijing University of Posts and Telecommunications
qinyanjun@bupt.edu.cn

Yuchen Fang*
Beijing University of Posts and Telecommunications
fangyuchen@bupt.edu.cn

Haiyong Luo†
Institute of Computing Technology, Chinese Academy of Sciences
yhluo@ict.ac.cn

Fang Zhao†
Beijing University of Posts and Telecommunications
zfsse@bupt.edu.cn

Chenxing Wang
Beijing University of Posts and Telecommunications
wangchenxing@bupt.edu.cn

ABSTRACT

Next Point-of-Interest (POI) recommendation is a pivotal issue for researchers in the field of location-based social networks. While many recent efforts show the effectiveness of recurrent neural network-based next POI recommendation algorithms, several important challenges have not been well addressed yet: (i) The majority of previous models only consider the dependence of consecutive visits, while ignoring the intricate dependencies of POIs in traces; (ii) The nature of hierarchical and the matching of sub-sequence in POI sequences are hardly model in prior methods; (iii) Most of the existing solutions neglect the interactions between two modals of POI and the density category. To tackle the above challenges, we propose an auto-correlation enhanced multi-modal Transformer network (AutoMTN) for the next POI recommendation. Particularly, AutoMTN uses the Transformer network to explicitly exploits connections of all the POIs along the trace. Besides, to discover the dependencies at the sub-sequence level and attend to cross-modal interactions between POI and category sequences, we replace self-attention in Transformer with the auto-correlation mechanism and design a multi-modal network. Experiments results on two real-world datasets demonstrate the ascendancy of AutoMTN contra state-of-the-art methods in the next POI recommendation.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Neural networks**.

KEYWORDS

Next POI recommendation; auto-correlation; multi-modal

*Equal contribution.
†Corresponding author.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

SIGIR '22, July 11–15, 2022, Madrid, Spain.
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8732-3/22/07.
https://doi.org/10.1145/3477495.3532033

ACM Reference Format:
Yanjun Qin, Yuchen Fang, Haiyong Luo, Fang Zhao, and Chenxing Wang. 2022. Next Point-of-Interest Recommendation with Auto-Correlation Enhanced Multi-Modal Transformer Network. In *Proceedings of the 45th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3477495.3532033

1 INTRODUCTION

The prompt expansion of location-based social networks (LBSNs), *e.g.*, Gowalla and Foursquare, promotes the research of point-of-interest (POI) recommender systems in both academia and industry [1, 13, 14]. The next POI recommendation is the most important POI recommendation task and aims to predict the possible future visited POIs by giving historical check-ins of users. It focuses more on exploiting the movement patterns and temporal influences hidden in the historical POI visited sequence of users compared with canonical POI recommendations.

The POI sequence in the traditional next POI recommendation algorithms is considered the Markov chain and the transition probability of POIs is integrated into the conventional matrix factorization models [2, 3]. Recently, deep learning-based methods have received substantial attention in the community. RNN and its variant LSTM have been broadly used among them. ST-RNN [6] as one of the earliest works on the next POI recommendation encoded the information of spatio-temporal distance in a recurrent network. The recent works Time-LSTM [18] and STGN [17] enhanced the performance of ST-RNN by taking the spatio-temporal gates as side information to capture the spatial and temporal preference of users. Besides, [8] proposes a variant of RNN to better capture the short-term sequential dependencies of POI sequences.

Though the RNN architecture-based methods recommend POIs in a data-driven manner, several key problems remain unsolved. 1) First, previous RNN-based methods are limited in the short-term contiguous visits, *i.e.*, they hardly model the implicit connections between visits far away on the timeline. For instance, a user consistently has dinner at a specific restaurant nearby the office on the workday evening, even if the user goes somewhere else during the workday. 2) Second, as shown in Figure 1, the same phase position in the green and red circles among periods naturally provides similar sub-visits, thus the POIs in the red circle on Day 2 can be

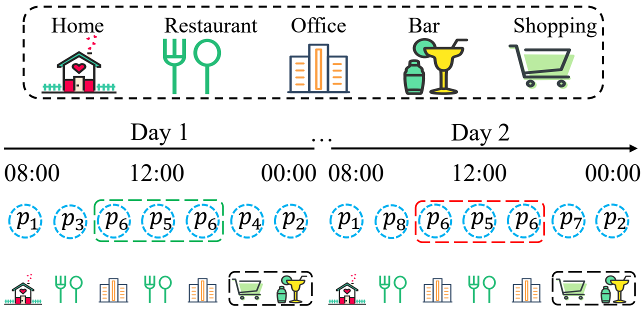


Figure 1: A toy example illustrating the trajectory of a user.

generated when we reveal a similar sub-sequence. However, prior works ignore the important influence of sub-sequence, which is beyond POI level sequential patterns. 3) The interaction between the category and location of POI is significant as the next location is influenced by the category, such as the user going to the same bar after shopping in Figure 1. However, the vast majority of existing algorithms fail to capture the cross-modal knowledge between POIs and categories.

Motivated by the above analysis, we propose an **auto**-correlation enhanced **m**ulti-modal **T**ransformer **n**etwork (AutoMTN) for next POI recommendation. Firstly, we consider utilizing the vanilla Transformer network in our paper to explicitly capture all connections in sequences at the POI level. However, this model fails to learn the information of category. To predict the next category of users to support the prediction of the next POI of users, we design a two-channel Transformer network to forecast POI and category, respectively. Besides, we replace the self-attention mechanism in the Transformer network with the sequence periodicity-based auto-correlation mechanism [10] to enhance our two-channel Transformer network, which can discover the dependencies and aggregate representation at the sub-sequence level. Finally, at the heart of our model is the directional cross-modal auto-correlation, which attends to interactions between POI and category sequences across distinct time steps and latently adapts information of sub-sequence from one modality to another. The main contributions of this paper are summarized as follows:

- To the best of our knowledge, our paper is the first to utilize the multi-modal Transformer network to gain the cross-modal information between POI and category sequences.
- We utilize the auto-correlation mechanism in our multi-modal Transformer network to learn the sub-sequence patterns.
- Experimental performances on two real-world LBSN datasets demonstrate that the proposed AutoMTN outperforms the state-of-the-art methods.

2 RELATED WORK

Most existing solutions are based on RNN and its variant LSTM. ST-RNN [6] directly adds the spatio-temporal distance information between successive POIs into a RNN architecture. Then, Time-LSTM [18] further integrates the time gate into the vanilla LSTM to capture the time information. STGN [17] further adds the spatial and temporal gates into the LSTM network to enhance the spatio-temporal representation power of LSTM. LSTPM [8] consists of not only a nonlocal network for learning long-term preferences but also a geo-dilated RNN network for modeling short-term preferences. Although these works have achieved satisfactory performance in the next POI recommendation, they only capture the correlations between consecutive POIs and fail to capture implicit relations at the sub-sequence level. Flashback [11] explicitly uses spatio-temporal contexts to find all hidden powerful states in history and ASPPA [16] proposes an adaptive sequence partitioner to automatically identify each semantic sub-sequence of POIs. Compared with them, our AutoMTN can capture the sub-sequence correlations but is not limited by the recurrent structure. Furthermore, the cross-modal information between the POI and category sequences can be gained in our model.

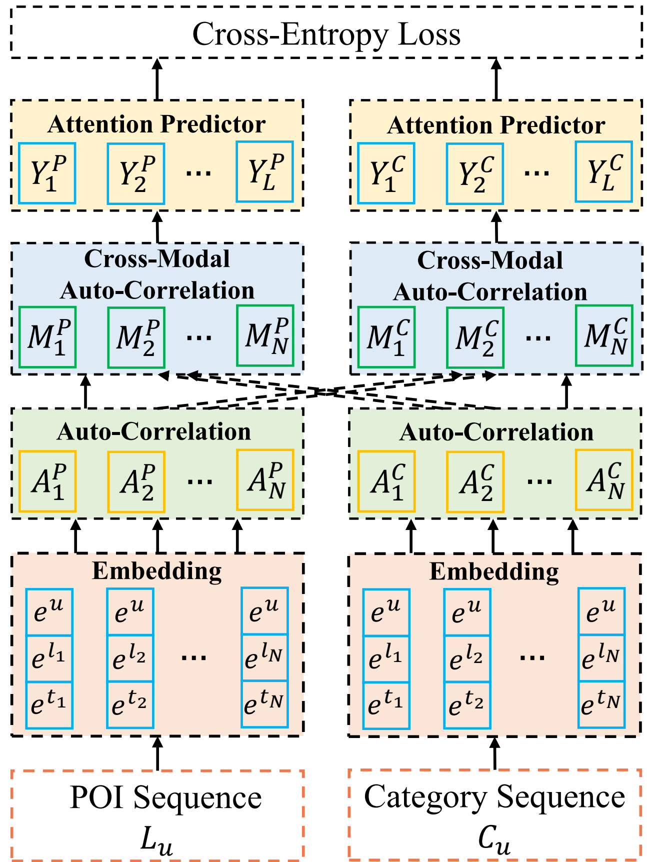


Figure 2: The overall architecture of AutoMTN.

3 PROBLEM FORMULATION

Let $U = \{u_1, u_2, ..., u_{|U|}\}$, $L = \{l_1, l_2, ..., l_{|L|}\}$, and $C = \{c_1, c_2, ..., c_{|C|}\}$ be the set of users, POIs, and categories, respectively. For each user $u \in U$, we can formulate the history POI sequence as $L_u = \{(u, l_t, \tau_t)|t = 1, 2, ..., N\}$, where each triple (u, l_t, τ_t) is the t-th visit of user $u \in U$ with POI $l_t \in L$ and time stamp τ_t. Similar with POI sequence, the correspond category sequence of user $u \in U$ can be

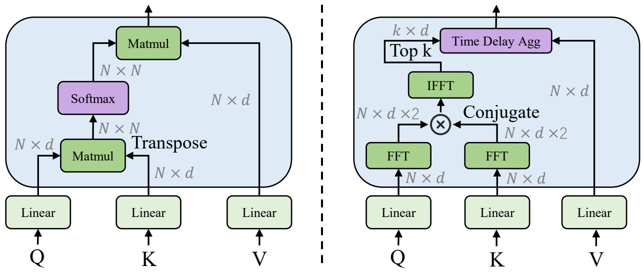


Figure 3: Self-attention (left) and auto-correlation (right).

formulated as $C_u = \{(u, c_t, \tau_t)|t = 1, 2, ..., N\}$. The problem of next POI recommendation is to recommend the top-K preferable POIs for user u after the history sequences, *i.e.*, we learn a personalized ranking function f_u for each user that maps the user's history POI and category sequences to a probability distribution for the target POI set. The function can be formulated as:

$$f_u = Pr(y \in L|L_u, C_u) \ . \tag{1}$$

4 METHODOLOGY

In this section, we introduce the details of AutoMTN. Figure 2 depicts the architecture, which consists of four modules including the embedding layer, the auto-correlation layer, the cross-modal auto-correlation layer, and the attention predictor. We further illustrate the auto-correlation mechanism in Figure 3.

4.1 Embedding Layer

The embedding layer is utilized to encode user, category, location, and time information into latent representations. For each user $u \in U$, POI $l \in L$, category $c \in C$, and time τ, we denote their embedded representations as $e^u \in \mathbb{R}^d$, $e^l \in \mathbb{R}^d$, $e^c \in \mathbb{R}^d$, $e^\tau \in \mathbb{R}^d$. The embedding is used in the other layers to convert the input into high-dimensional vectors to enhance the representation power of our model. Here, temporal features are represented by the time slot, where we have 24 slots of hours in a day. Therefore, each feature in our work has a unique ID. The input dimensions of user, POI, category, and time are $|U|$, $|L|$, $|C|$, and 24. The embedding output of the triple in the POI and category sequence are the sum $e^P = e^u + e^l + e^\tau \in \mathbb{R}^d$ and $e^C = e^u + e^c + e^\tau \in \mathbb{R}^d$. The embedding of each POI and category sequence are denoted as $E(L_u) = \{e_1^P, e_2^P, ..., e_N^P\} \in \mathbb{R}^{N \times d}$ and $E(C_u) = \{e_1^C, e_2^C, ..., e_N^C\} \in \mathbb{R}^{N \times d}$.

4.2 Auto-Correlation Layer

Previous works are mainly based on the recurrent structure, which always neglects the information with non-consecutive visits. Although the well-known self-attention can capture all the point-to-point interactions within the sequence, which fails to extract the correlations at the sub-sequence level. Therefore, we leverage the auto-correlation mechanism to expand the information utilization by sub-sequence-wise relationships. Specifically, auto-correlation [10] discovers the dependencies of sub-sequences by calculating their autocorrelation and aggregates similar sub-sequences through the time delay aggregation.

Dependencies of Sub-sequences. The autocorrelation can reflect the time-delay pattern between a POI sequence and its ϵ lag sequence. In detail, the autocorrelation $\mathcal{R}_{\mathcal{X}\mathcal{X}}(\epsilon)$ of a real discrete-time process $\{\mathcal{X}_t\}$ can be denoted as the following equation according to the stochastic process theory [7].

$$\mathcal{R}_{\mathcal{X}\mathcal{X}}(\epsilon) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{X}_t \mathcal{X}_{t-\epsilon} \ . \tag{2}$$

Eq. (2) can be optimized to $O(logN)$ complexity by the Fast Fourier Transforms (FFT) based on the Wiener–Khinchin theorem [9].

Time Delay Aggregation. We choose k lag lengths with the highest autocorrelation from the calculated autocorrelation of all lag lengths and then align these selected sub-sequences. Finally, we aggregate information from selected sub-sequences. The aggregation can be formalized as:

$$\begin{aligned} \epsilon_1, ..., \epsilon_k &= \underset{\epsilon \in \{1,...,N\}}{argTopk} \left(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\epsilon)\right) \\ \tilde{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\epsilon_1), ..., \tilde{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\epsilon_k) &= SM(\mathcal{R}_{\mathcal{Q},\mathcal{K}}(\epsilon_1), ..., \mathcal{R}_{\mathcal{Q},\mathcal{K}}(\epsilon_k)) \\ AC(\mathcal{Q}, \mathcal{K}, \mathcal{V}) &= \sum_{i=1}^{k} Roll(\mathcal{V}, \epsilon_i) \tilde{\mathcal{R}}_{\mathcal{Q},\mathcal{K}}(\epsilon_i) \ , \end{aligned} \tag{3}$$

where $AC(\cdot)$ and $SM(\cdot)$ denote the auto-correlation mechanism and softmax normalization. k in our paper is set to $logN$ to keep the $O(NlogN)$ complexity. $Roll(\mathcal{X}, \epsilon)$ operation can align the $\mathcal{X}$ with time delay ϵ, *i.e.*, the first ϵ elements in the time delay sequence are re-introduced at the last index.

For the embedding of POI and category sequence, the auto-correlation mechanism can be formulated as:

$$\begin{aligned} A^P &= AC(E(L_u)W^{Q^{P_1}}, E(L_u)W^{K^{P_1}}, E(L_u)W^{V^{P_1}}) \\ A^C &= AC(E(C_u)W^{Q^{C_1}}, E(C_u)W^{K^{C_1}}, E(C_u)W^{V^{C_1}}) \ , \end{aligned} \tag{4}$$

where $W^{Q^{P_1}}, W^{K^{P_1}}, W^{V^{P_1}}, W^{Q^{C_1}}, W^{K^{C_1}}, W^{V^{C_1}} \in \mathbb{R}^{d \times d}$ denote the parameter matrices of projections. $A^P, A^C \in \mathbb{R}^{N \times d}$ are the output of POI and category sequences of this layer.

4.3 Cross-Modal Auto-Correlation Layer

Different from previous methods that ignore or parallel computing category knowledge, we design the cross-modal auto-correlation that enables one modality for receiving information from another modality, *i.e.*, POI and category sequences to obtain auxiliary information from each other. Specifically, we treat the POI sequence as the query in the auto-correlation to extract the side information of the category at the sub-sequence level, and vice versa. The cross-modal auto-correlation can be formulated as:

$$\begin{aligned} M^P &= AC(A^P W^{Q^{P_2}}, A^C W^{K^{C_2}}, A^C W^{V^{C_2}}) \\ M^C &= AC(A^C W^{Q^{C_2}}, A^P W^{K^{P_2}}, A^P W^{V^{P_2}}) \ , \end{aligned} \tag{5}$$

where $W^{Q^{P_2}}, W^{K^{P_2}}, W^{V^{P_2}}, W^{Q^{C_2}}, W^{K^{C_2}}, W^{V^{C_2}} \in \mathbb{R}^{d \times d}$ denote the parameter matrices of projections. $M^P, M^C \in \mathbb{R}^{N \times d}$ are the output of POI and category sequences of this layer.

4.4 Attention Predictor

We compute the probability of a candidate POI and category based on the learned representations of $M^P, M^C \in \mathbb{R}^{N \times d}$, the embedded

Table 1: Performance comparison on two datasets. Bold: Best, underline: Second best.

		NYC				TKY			
		Acc@5	Acc@10	MRR@5	MRR@10	Acc@5	Acc@10	MRR@5	MRR@10
Baselines	RNN	0.2814	0.3233	0.1914	0.1970	0.2465	0.2944	0.1714	0.1778
	LSTM	0.3134	0.3651	0.2066	0.2136	0.2634	0.3125	0.1799	0.1865
	STGN	0.3400	<u>0.3879</u>	<u>0.2327</u>	<u>0.2393</u>	0.2786	0.3284	0.1908	0.1975
	LSTPM	0.3326	0.3778	0.2289	0.2326	0.2658	0.3207	0.1855	0.1920
	Flashback	<u>0.3448</u>	0.3877	0.2302	0.2381	<u>0.3353</u>	<u>0.4000</u>	<u>0.2253</u>	<u>0.2340</u>
Ours	MTN	0.4190	0.5066	0.2636	0.2755	0.3697	0.4587	0.2310	0.2491
	AutoTN	0.4063	0.4933	0.2599	0.2689	0.3504	0.4325	0.2297	0.2409
	AutoMTN	**0.4370**	**0.5289**	**0.2742**	**0.2867**	**0.3802**	**0.4687**	**0.2434**	**0.2553**
	Improvement	26.74%	36.35%	17.83%	19.81%	13.39%	17.18%	8.03%	9.10%

Table 2: Statistics of NYC and TKY datasets.

Dataset	#User	#POI	#Category	#Check-in
NYC	866	7725	345	171753
TKY	2034	11650	313	482677

POI candidates $E(L) = \{e^l_1, ..., e^l_{|L|}\} \in \mathbb{R}^{|L|\times d}$, the embedded category candidates $E(C) = \{e^c_1, ..., e^c_{|C|}\} \in \mathbb{R}^{|C|\times d}$, and the distance matrix $D^{N\times|L|}$ between POIs in input sequence and candidates. The attention predictor can be formulated as:

$$\hat{Y}^P = S(SM(\frac{E(L)M^{P^T} + D^T}{\sqrt{d}})) , \hat{Y}^C = S(SM(\frac{E(C)M^{C^T}}{\sqrt{d}})) , \quad (6)$$

where $S(\cdot)$ is a weighted sum of the last dimension.

Finally, We apply the cross-entropy loss function to quantify the discrepancies of predicted values and ground truths:

$$\mathcal{L} = -\frac{1}{B}\sum_{b=1}^{B} y_b^{P^T} log(\hat{y}_b^P) + \eta y_b^{C^T} log(\hat{y}_b^C) \quad (7)$$

where b is the index of training samples, y_b^P and y_b^C are the one-hot vectors of POI and category ground truths respectively. η indicates the weight hyperparameter of the next category loss.

5 EXPERIMENTS AND RESULTS

5.1 Experimental Settings

Datasets and Metrics. We evaluate AutoMTN on two public real-world LBSN datasets[1], they were collected from New York City (NYC) and Tokyo (TKY) and have been widely used for the next POI recommendation [5, 12]. Both two datasets include the world-wide POI data of users from April 2012 to February 2013. Descriptive statistics for those datasets are shown in Table 2. We filter out POIs in datasets that are visited by less than 5 times, and we only maintain the recent 80 POIs visited for each user. We chronologically split the check-in history of each user into 60% for train, 20% for validation, and 20% for testing. Two standard metrics are adopted in our paper to test the performance of all models: average Accuracy@N (Acc@K) and Mean Reciprocal Rank@K (MRR@K), where $K = 5, 10$.

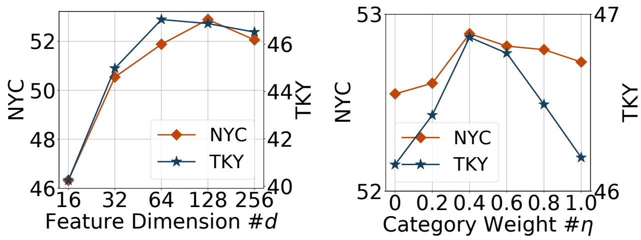
Baselines. We compare AutoMTN with five models: 1) RNN [15]: This method captures the temporal dependence from check-in sequences through a standard recurrent structure; 2) LSTM [4]: A popular variant of RNN, which is a gated recurrent structure; 3) STGN [17]: A variant of LSTM, which adds additional gates controlled by the spatio-temporal distances in LSTM; 4) LSTPM [8]: This method uses a nonlocal network for learning long-term preferences and a geo-dilated RNN network for modeling short-term preferences; 5) Flashback [11]: This method captures temporal dependence through a general RNN architecture and uses spatio-temporal contexts to find hidden states with high predictive power in the past.

Implementation Details. AutoMTN is implemented using PyTorch with Tesla A100. For consistency, we apply the same dimension size d for all embeddings and weight matrices in this paper. Specifically, we set $d = 128$ and the category loss weight η to 0.4. All the trainable parameters in our model are optimized by the Adam optimizer with a batch size of 32 and a learning rate of 0.001.

5.2 Results and Analysis

Performance Comparison. Table 1 shows the evaluation results of all models on the next POI recommendation task on both NYC and TKY datasets. From the statistics in the table, we can draw the following observations: 1) The results on both two datasets demonstrate that the proposed AutoMTN significantly outperforms all baselines on Acc and MRR evaluation metrics. On the NYC dataset, compared with the second-best methods, AutoMTN gains 26.74%, 36.35%, 17.83%, and 19.81% improvement on Acc@5, Acc@10, MRR@5, and MRR@10, respectively. On the TKY dataset, AutoMTN acquires around 10% advancement on average over the best-performing baselines, which obviously shows the effectiveness of AutoMTN; 2) STGN and Flashback perform better than RNN, LSTM, and LSTPM. This is because STGN adds extra spatio-temporal prior knowledge and Flashback aggregates information from all historical POIs. However, due to the lack of capturing information from the density cross-modal category and sub-sequence, STGN and Flashback result in sub-optimal performance.

[1] http://www-public.it-sudparis.eu/~zhang_da/pub/dataset_tsmc2014.zip



(a) (b)

Figure 4: Parameter sensitivity analysis (Acc@10).

Ablation Study. To verify the contribution of each proposed component in AutoMTN, we conduct ablation experiments with two variants of AutoMTN that remove one component at a time. The first variant, MTN replaces the auto-correlation in our model with the vanilla self-attention. The second variant, AutoTN removes the channel of category in our model. Table 1 shows the performance of ablation tests on NYC and TKY datasets. When the channel of the category is removed, AutoTN has a clear performance drop proving that gaining category information and learning the interaction between location and category is effective for alleviating the sparsity of POI. The matching of sub-sequence is essential for the next POI task because of the performance degradation of MTN.

Hyperparameter Analysis. We further investigate the influence of different hyperparameter settings to AutoMTN, the performance (Acc@10) are depicted in Figure 4. In Figure 4a, the results of AutoMTN climb up as the dimension of embeddings and weights increases and gradually becomes stable with the size 128. Figure 4b illustrates the performance of varying the category loss weights η, which control the importance of the category prediction tasks. We find that AutoMTN performs best when $\eta = 0.4$.

6 CONCLUSION AND FUTURE WORK

In this paper, we propose an auto-correlation enhanced multi-modal Transformer network for the next POI recommendation, which is a two-channel Transformer architecture. In particular, we replace the self-attention with the auto-correlation in each channel to discover the sequential dependencies and aggregate information at the sub-sequence level. Moreover, we devise the cross-modal auto-correlation layer, which latently adapts the information of sub-sequence from one modality to another. Experimental results on two LBSN datasets show the effectiveness of our AutoMTN.

ACKNOWLEDGEMENT

This work is supported by the Joint Research Fund for Beijing Natural Science Foundation and Haidian Original Innovation under grant L192004, the Beijing Natural Science Foundation under grant 4212024 and 4222034, and the Beijing Key Laboratory of Mobile Computing and Pervasive Device.

REFERENCES

[1] Chen Cheng, Haiqin Yang, Michael R Lyu, and Irwin King. 2013. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, August 3–9, 2013, Beijing, China.* AAAI Press, California, USA, 2605–2611. https://doi.org/10.5555/2540128.2540504

[2] Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. 2015. Personalized ranking metric embedding for next new poi recommendation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, July 25–31, 2015, Buenos Aires, Argentina.* AAAI Press, California, USA, 2069–2075. https://doi.org/10.5555/2832415.2832536

[3] Jing He, Xin Li, Lejian Liao, Dandan Song, and William Cheung. 2016. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12–17, 2016, Arizona, USA.* AAAI Press, California, USA, 137–143. https://doi.org/10.5555/3015812.3015833

[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[5] Dongliang Liao, Weiqing Liu, Yuan Zhong, Jing Li, and Guowei Wang. 2018. Predicting Activity and Location with Multi-task Context Aware Recurrent Neural Network.. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, July 13-19, 2018, Stockholm, Sweden.* IJCAI, Vienna, Austria, 3435–3441. https://doi.org/10.24963/ijcai.2018/477

[6] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the Thirtieth AAAI conference on artificial intelligence, February 12-17, 2016, Arizona, USA.* AAAI Press, California, USA, 194–200. https://doi.org/10.5555/3015812.3015841

[7] Athanasios Papoulis. 1965. *Random variables and stochastic processes.* McGraw Hill.

[8] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2020. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, USA.* AAAI Press, California, USA, 214–221. https://doi.org/10.1609/aaai.v34i01.5353

[9] Norbert Wiener. 1930. Generalized harmonic analysis. *Acta mathematica* 55, 1 (1930), 117–258.

[10] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems, December 6-14, 2021, Virtual.*

[11] Dingqi Yang, Benjamin Fankhauser, Paolo Rosso, and Philippe Cudre-Mauroux. 2020. Location prediction over sparse user mobility traces using RNNs: Flashback in hidden states!. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, January, 2021, Virtual.* IJCAI, Vienna, Austria, 2184–2190. https://doi.org/10.5555/3491440.3491742

[12] Dingqi Yang, Daqing Zhang, Vincent W Zheng, and Zhiyong Yu. 2014. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 1 (2014), 129–142.

[13] Hongzhi Yin and Bin Cui. 2016. *Spatio-temporal recommendation in social media.* Springer.

[14] Wei Zhang and Jianyong Wang. 2015. Location and time aware social collaborative retrieval for new successive point-of-interest recommendation. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Oct 18-23, 2015, Melbourne, Australia.* Association for Computing Machinery, New York, USA, 1221–1230. https://doi.org/10.1145/2806416.2806564

[15] Yuyu Zhang, Hanjun Dai, Chang Xu, Jun Feng, Taifeng Wang, Jiang Bian, Bin Wang, and Tie-Yan Liu. 2014. Sequential click prediction for sponsored search with recurrent neural networks. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27–31, 2014, Québec, Canada.* AAAI Press, California, USA, 1369–1375. https://doi.org/10.5555/2893873.2894086

[16] Kangzhi Zhao, Yong Zhang, Hongzhi Yin, Jin Wang, Kai Zheng, Xiaofang Zhou, and Chunxiao Xing. 2020. Discovering Subsequence Patterns for Next POI Recommendation.. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, January, 2021, Virtual.* IJCAI, Vienna, Austria, 3216–3222. https://doi.org/10.5555/3491440.3491885

[17] Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Jiajie Xu, Zhixu Li, Fuzhen Zhuang, Victor S Sheng, and Xiaofang Zhou. 2019. Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, January 27-February 1, 2019, Hawaii, USA.* AAAI Press, California, USA, 5877–5884. https://doi.org/10.1609/aaai.v33i01.33015877

[18] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. 2017. What to Do Next: Modeling User Behaviors by Time-LSTM.. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, August 19-25, 2017, Melbourne, Australia.* IJCAI, Vienna, Austria, 3602–3608. https://doi.org/10.5555/3172077.3172393