Meta-learning Enhanced Next POI Recommendation by Leveraging Check-ins from Auxiliary Cities

By

Jinze Wang

A thesis submitted to Macquarie University
for the degree of Masters of Research
Faculty of Science and Engineering
Department of Computing
August 2022

Examiner's Copy

Typeset in $\LaTeX\,2_\varepsilon$.

This thesis is submitted to Macquarie University in fulfilment of the requirement for the Degree of Master of Research.

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Jinze Wang

Declaration

This thesis contains work that has not been submitted previously, in whole or in part, for any other academic award and is solely my original research, except where acknowledged.

This work has been carried out since July 2021, under the supervision of Dr. Zhu Sun and Prof. Yan Wang.

Acknowledgements

I would like to take this opportunity to convey my special thanks to my supervisors Dr. Zhu Sun and Prof. Yan Wang who encouraged me to do research. This thesis would not have been possible without their incredible support and superb energy. I would also like to thank Lu Zhang for her useful comments and endless discussions whenever I was stuck.

Abstract

Next point-of-interest (POI) recommendation has attracted a considerate amount of attention in the area of research recently to recommend the next POI where users are most likely to visit at the next time step. However, most existing next POI recommendation algorithms suffer from severe data sparsity issues, due to the scarcity of historical check-in data. Existing studies mainly resort to side information, such as POI categories, to mitigate the data sparsity problem, but ignores the rich check-in information from other cities. To this end, we explore how knowledge transfer from data-rich cities with diverse user patterns can help improve the next POI recommendation performance for cities with sparse check-ins. Accordingly, we propose a novel Meta-learning Enhanced next POI Recommendation (MERec) framework by leveraging check-in data from auxiliary cities, which incorporates the correlation of check-in behaviors among cities into the meta-learning paradigm. Concretely, the MERec framework takes into account the user check-in patterns of the target and auxiliary cities in terms of culture, urban structure, resident behaviour, etc., and transfers more relevant knowledge from more correlated cities. Extensive experiments on four real-world datasets demonstrate the superiority of our proposed MERec framework.

Contents

List of Figures

List of Tables

1 Introduction

The next POI recommendation [Islam et al. (2020)], which is to recommend a user a location where they are most likely to go at a specific time in the upcoming hours, benefits many location-based companies and individuals. However, the data in many cities is extremely sparse due to the limited number of user-POI interactions, which is a major challenge for the next POI recommendation tasks. Table 1.1 shows the number of user-POI interactions for four different cities on Foursquare[1], where we can clearly see that the user check-in data density in some cities are extremely low, such as Singapore with density being 0.05%. Consequently, with only the historical check-in data, it is not possible to train comparable models for these data-insufficient cities.

To this end, most of the current research is devoted to augmenting the check-in

[1] https://foursquare.com/

TABLE 1.1: Statistics of four datasets from Foursquare.

	#User	#POI	#Check-in	#Category	Density
Calgary	435	3,013	13,911	293	**1.06%**
Phoenix	2,945	7,247	47,980	344	**0.22%**
Singapore	8,648	33,712	355,337	398	**0.12%**
New York	16,387	56,252	511,431	420	**0.05%**

data of cities with side information, e.g., POI category, to alleviate the data sparsity issue. They are built upon various techniques, ranging from the simple matrix decomposition [Lian et al. (2014); Wang et al. (2021)], Markov chain models [Cheng et al. (2013)], to advanced deep learning frameworks, e.g., graph neural networks [Qian et al. (2019); Xie et al. (2016)] and recurrent neural networks [Huang et al. (2019); Zhang et al. (2021); Liu et al. (2021)]. Despite great success of those methods, they heavily rely on sufficient training data, thus merely achieving limited improvements because of the restriction of data sparsity issue.

To ease this shortcoming, we thus conduct in-depth data analysis on the check-in data across different cities in Chapter 3, where we surprisingly find out that the check-in behaviors among different cities may share certain common patterns. That is, there are overlapping behavioral patterns across different cities, for instance, the transition pattern *Drink→Travel&Transport* is shared by the four cities in Table 1.1, which suggests that users of different cities may always take public transport home after drinking. This, consequently, inspires us to leverage data-rich cities to facilitate data-sparse cities, so as to further improve the performance of the next POI recommendations. On the other hand, non-overlapping behavioural patterns can be also noted from city to city despite of the shared patterns due to the inherent diversity of culture, structure and geographical location of cities [Chen et al. (2021); Tan et al. (2021)]. For instance, the transition pattern *Travel&Transport→Shop* is quite common in Singapore due to the convenient public transportation, while rare in the other three cities. In this sense, blindly leveraging all check-ins in the data-rich city to augment the data-sparse city may inversely hurt the recommendation accuracy.

Therefore, we are facing with two major challenges when transferring knowledge from data-rich cities to augment data-sparse cities for performance-enhanced next POI recommendation.

- *What to Transfer.* As observed from the data, different cities do not have overlapping POIs, making the knowledge (i.e., check-ins) transfer challenging. In contrast to the context of e-commerce, overlapping items can be found on shopping sites in different regions [Bonab et al. (2021)]. In our study, there is no intersection of POIs from different cities. Furthermore, according to the analysis of the dataset [Chen et al. (2021)], most user's next POI is within the city, which greatly reduces the interaction of POIs across cities.

- *How to Transfer.* Cultural and structure and geographical diversity of cities make users have different check-in patterns among cities, while only the common patterns may help enhance the next POI recommendation accuracy. In other words, although data-rich cities can be used as a source city to enrich the target data-sparse city, different cities differ greatly in the distribution regarding user check-in patterns. Simply transferring the entire data from data-rich cities may not bring satisfying recommendation results. As a result, it inevitably increases the difficulty in developing suitable transfer algorithms.

Although some existing methods exploit transfer-based approaches to enhance the target cities [Ding et al. (2019); Zhang et al. (2020a)] for better next POI recommendation, the results are not satisfying due to the limited number of overlapping POIs among cities. For example, tea-houses that are popular in Asian cities are hard to find in American cities. Besides, there are also some meta-learning based approaches that adopt knowledge transfer to provide potential solutions to address the above challenges [Chen et al. (2021); Tan et al. (2021); Cui et al. (2021)]. Nevertheless, they directly ingore the difference of user behavior patterns across cities.

In this thesis, to supply more precise next POI recommendation via addressing the above challenges, we introduce a novel Mete-learning Enhanced next POI Recommendation (MERec) framework by leveraging check-in data from auxiliary cities to augment target cities. Most importantly, it delicately takes into account the correlation of behavioral patterns across different cites into the meta-learning paradigm by sticking to "paying more attention to more correlated knowledge". Specifically, MERec is mainly composed of two components, including a two-channel encoder (i.e., category- and POI-level encoders) and a city-specific decoder. Firstly, the category-level encoder consists of the meta-leaning paradigm and long-short term networks (LSTM), and is designed to obtain expressive representations of categories by capturing the shared category transition patterns across different cities. Secondly, the POI-level encoder aims to learn accurate representations of POIs in target cities through LSTM. Lastly, the city-specific decoder aggregates the latent representations of two channels to perform next POI prediction task on the target city.

In summary, the main contributions of this thesis resides in three fields.

- We propose a novel Mete-learning Enhanced next POI Recommendation (MERec) framework by leveraging user check-in data from auxiliary cities to augment target cities, thus alleviating the data sparsity issue.

- Holding the principle "paying more attention to more correlated knowledge", MERec transfers more relevant knowledge from more correlated cities, i.e., transferring more category-level check-in patterns (what to transfer) from more correlated auxiliary cities (how to transfer) to the target city.

- We conduct extensive experiments on four real-world datasets to validate the effectiveness of our proposed MERec. The experimental results demonstrate the superiority of MERec against state of-the-art (SOTA) baselines.

2

Related Work

This chapter briefly reviews the related work to next POI recommendation, including general POI recommendation, next POI recommendation, transfer learning based next POI recommendation as well as meta-learning based next POI recommendation.

2.1 General POI Recommendation

Recently, more sophisticated approaches have been introduced to leverage additional information for POI recommendation [Adams et al. (2010); Gu et al. (2010)], such as social influence, geographical information, temporal information, review information and transition between POIs. For instance, a topic model is proposed by [Kurashima et al. (2013)], a POI is sampled according to the topics and distances of historical POIs accessed by the target user. Levandoski et al. (2012) applied an item-based

CF model for POI recommendation, taking into account a travel penalty, which is proportional to the distance between the POI and the target user. Ye et al. (2010, 2011) models geographic influence through a Bayesian CF model in the framework of a user-based collaborative filtering (CF) model, and also takes into account social influence. Liu et al. (2013) approximates the geographical correlations of check-in POIs by a power-law distribution. Zhang and Chow (2013) directly carries out kernel density estimation for this distribution. Later, more comprehensive information is considered by [Ye et al. (2010); Cheng et al. (2012)], such as the multi-center of user check-in patterns, and the skewed user check-in frequency. Moreover, time preference is introduced to boost the efficiency and effectiveness of POI recommendations [Yuan et al. (2013); Gao et al. (2013)]. Lian et al. (2014) factorized geographic information into the weighting matrix to boost the effectiveness of POI recommendations. Besides, Liu et al. (2016b) designed a bi-weighted low-rank graph construction model that combines users' interests and their shifting sequential preferences with time interval evaluation to provide time-specific POI recommendations.

2.2 Next POI Recommendation

The next POI recommendation is an emerging challenge that is more challenging than the general POI recommendation. Early studies usually employ matrix factorization models to characterize the personalized sequential patterns of users. For instance, Zhao et al. (2016) proposed a pairwise tensor factorization technique (STELLAR) for next POI recommendation, a ranking-based framework that can incorporate fine-grained temporal context. A personalized ranking metric embedding (PRME) approach was presented to reflect user preferences and POI sequential transitions [Feng et al. (2015)]. Meanwhile, Markov chain models have also been used to model the sequential influence. For example, a personalized Markov chain model with factorization recommends continuous POIs to the target user [Cheng et al. (2013)]. Similarly, an additive Markov chain model was developed for predicting the probability of continuous transitivity [Zhang et al. (2014)]. Besides, a hybrid Hidden Markov Model is proposed to learn the delivery pattern of POI categories for successive user check-ins [Ye et al. (2013)].

Recently, Recurrent neural networks (RNNs) such as long short-term memory (LSTM) [Hochreiter and Schmidhuber (1997)] have showed breakthrough capability in modeling sequential check-in behavior for the next POI recommendation. Since RNN can tackle sequentially ordered data very well, existing works mainly concentrated on leveraging users' sequential preference on POIs by incorporating diverse context information into RNNs framework. For instance, Liu et al. (2016a) proposed Recurrent Neural Networks based on Spatial Temporal (ST-RNN) model to capture the periodical spatial and temporal contexts. The RNN was used to learn to produce new user paths for next stop-over prediction by simulating temporal correlations between POI categories [Palumbo et al. (2017)]. Zhang et al. (2021) devised a LSTM model using a two-channel encoder and a task-specific decoder for catching the sequential correlations of activities and location preferences based on side information, e.g., POIs category information.

Despite of the great success of these methods, most of them rely on sufficient training data and require extensive types of side information. As a result, the performance improvements are heavily restricted by the severe data sparsity issue. In this sense, transferring knowledge from data-rich cities to data-sparse cities becomes necessary to further help boost the performance of the next POI recommendation.

2.3 Transfer Learning for Next POI Recommendation

Transfer learning primarily concerns transferring knowledge from the source domain to the target, so as to resolve the data sparsity issue in the target domain [Farseev et al. (2017)]. A collaborative filtering model is proposed by Wang et al. (2018) to merge data from different sources by using neighborhood information of common users/items. Man et al. (2017) introduced an embedding and mapping framework for cross-domain recommendation, which learns mapping functions through latent vector projections of different domains. However, for next POI recommendations, transfer learning inherently suffer from several limitations. First, the direct transfer of knowledge from auxiliary cities to the target city ignores the fact that the structure and user behavior patterns of different cities produce different data distributions. Second, as suggested by [Chen et al. (2021)], transfer learning requires a sufficient number of overlapping users and or items between the source and target domains for the knowledge to be transferred.

2.4 Meta-learning for Next POI Recommendation

Inspired by human learning from previous relevant tasks to quickly learn new skills, meta-learning [Vanschoren (2018)], designing to transfer the knowledge learned from multiple tasks to efficiently accomplish different new tasks, has achieve significant success mainly in few-show learning applications. The four popular approaches are listed below: 1) learning a proper initialization from which the model parameters can be updated within a couple of gradients steps [Yao et al. (2019)]. 2) learning a valid distance metric between instances [Snell et al. (2017)]; 3) learning a meta-optimizer which can rapidly optimize the model parameters [Ravi and Larochelle (2016)]; and 4) using a recurrent neural network equipped with either external or internal memory storing and querying meta-knowledge [Mishra et al. (2017); Munkhdalai et al. (2018)];

Nevertheless, only a few attempts have been introduced to alleviate the data sparsity issue for next POI recommendation. For instance, Curriculum Hardness Aware Meta-Learning (CHAML) framework was proposed by Chen et al. (2021), which takes into account a city-level curriculum and city- and user-level hardness in meta training. Tan et al. (2021) introduced a meta-learning enhanced neural ordinary differential equation (ODE) method, which models city-irrelevant information and city side information to achieve citywide next POI recommendation. Cui et al. (2021) proposed a meta-learned sequential-knowledge-aware recommender (Meta-SKR), which utilizes sequential, spatio-temporal, and social knowledge to recommend the next POI for users.

Unfortunately, the above meta-learning based next POI recommendation methods completely ignore the diversity of user check-in patterns across different cities. To this end, this thesis proposes a novel meta-learning enhanced approach – MERec by leveraging check-ins from auxiliary cities to augment the target city. By holding the principle "paying more attention to more correlated knowledge", the proposed MERec is capable of further boosting the performance of next POI recommendation via better resolving the data sparsity issue.

3

Data Collection and Analysis

In this chapter, we first describe in detail the data collection process and provide statistics for the collected data. Then we conduct in-depth analysis to gain important observations to guide our model design.

3.1 Data Collection

We collect four datasets, i.e., Calgary (CAL), Phoenix (PHO), Singapore (SIN), New York (NYC), from Foursquare [Yang et al. (2016)], which are widely-used datasets in the next POI recommendation Zhang et al. (2020b, 2021). It is worth noting that the behavioural patterns of users can vary considerably from city to city, depending on the urban structure and culture of the city. Our goal is to transfer relevant knowledge from different cities to alleviate the data sparsity problem in the target cities. To make the selection of cities more reasonable, we first selected New York and Phoenix as the two cities from the USA. New York has a complex urban structure and a large amount of check-in data, while Phoenix is not as rich in the number and type of POIs as New York. Secondly, we chose Calgary in Canada, as the third city. The reason for this is that although Canada and the two USA cities are located at North America, they possess slightly different cultures; besides Calgary has a much simpler city structure than Phoenix. Lastly, we choose Singapore as the fourth city. This is mainly because Singapore is on a different continent to the previous three cities, and there are clear differences in culture and urban structure among them. Meanwhile, Singapore has more POIs and categories than Phoenix, and is a suitable auxiliary city thanks to the diversity of user behaviour patterns.

Following [Zhang et al. (2021, 2020b)], each check-in is formed as (u, p, t, c, g) meaning that user u visits POI p at time t, where p is associated with category c as well as geocoded by g (latitude and longitude of p). For each user, we order his check-in records via the timestamp information, and then divide them into different sequences by day. The statistics of the four datasets are shown in Table 1.1.

3.2 Data Analysis

Our goal in this section is to analyze the similarities and demonstrate differences regarding user behavioral patterns across cities, which could better guide the knowledge transfer process from the auxiliary cities to the target cities for a more accurate next POI recommendation. However, this is non-trivial due to the non-overlapping POIs across different cities. Fortunately, the POIs in the four cities share the same set of categories, which consequently inspires us to analyze both the distribution of POIs and user behavioural patterns at the category level in different cities, so as to better uncover the similarities and differences of the patterns.

Distribution of POIs at Category Level. The number of POIs under each category varies significantly from city to city due to differences in geographical location, urban structure culture. To better understand user behavior patterns, we first investigate the nature of the POI distribution in each city. According to [Zhang et al. (2020b); Sun et al. (2021)], the POIs in the four cities are characterized by 10 first-level categories, including Arts & Entertainment (AE), College & University (CU), Drink (DR), Food (FO), Nightlife Spot (NS), Outdoor & Recreation (OR), Professional & Other Places (PO), Residence (RE), Shop & Service (SS), Travel & Transport (TT).

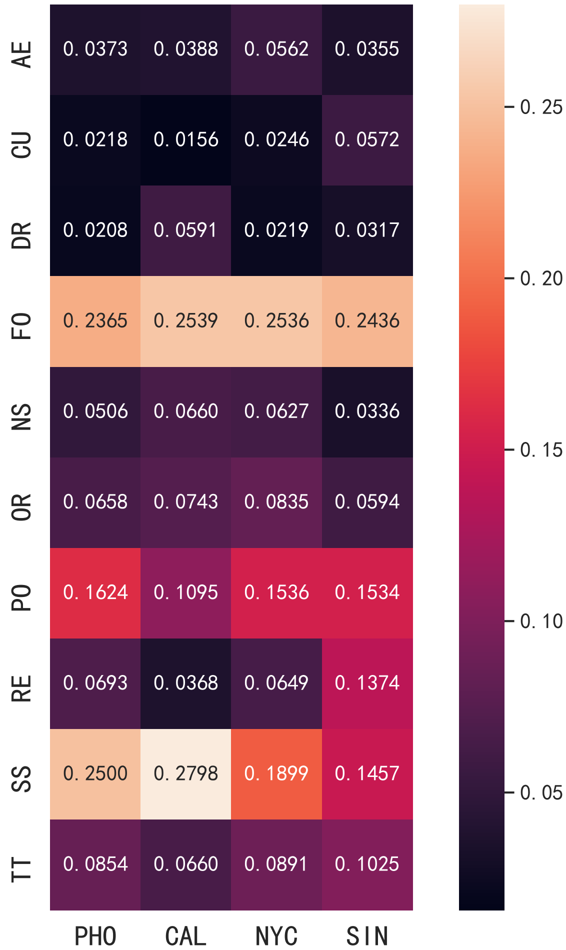


Figure 3.1: The POI distribution at category level across the four cities.

Fig. 3.1 depicts the POI distribution at category level across the four cities. We can clearly note that there are certain similarities between the POI distribution of the four cities, while at the same time significant differences exist to some extent. Firstly, the proportion of Food-related POIs is quite high across the four cities. In particular, New York and Singapore have the highest number of food-related POIs. Regarding Phoenix and Calgary have the second highest number of food-related POIs, whereas Shop & Service ranks first regarding the number of POIs in the two cities. In contrast, New York and Singapore have slightly lower numbers of Shop & Service. Professional & Other Places is the second most popular type of POIs in Singapore, but is far less represented in Calgary than in the other cities. Secondly, in terms of College & University, the distribution of Phoenix and New York are relatively similar, while Calgary has the lowest proportion of College in the city. Unlike the other three cities, Singapore has a much higher ratio of College & University.

Correlation of POIs at Category Level. Based on the above distribution, both similarity and dissimilarity can be observed across different cities. The main goal of our study is to transfer check-in data from more correlated cities to assist the target cities by holding the principle of "paying more attention to more correlated knowledge". Therefore, it is essential to analyze the correlation between cities. To start, we first analyse the correlation of POIs at category level. Given two cities, $A = [A_1, A_2...A_i]$ and $B = [B_1, B_2...B_i]$ denote the number of POIs under each category for per city. Accordingly, the correlation γ_{cor} can be calculated as below:

$$\gamma_{cor} = \frac{\sum(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{(A_i - \bar{A})^2}\sqrt{(B_i - \bar{B})^2}}, \tag{3.1}$$

where A_i and B_i are the number of POIs under category c_i in city A and city B, respectively.

By using Eq.(3.1), we can figure out the correlation of POIs at category-level among the four cities, which is illustrated in Fig. 3.2. Interestingly, we observe the highest correlation (i.e., 0.9665) is possessed by New York and Phoenix. This suggests that cities in the same country (United States) have a higher correlation in terms of urban structure and culture compared to cities in other countries. In addition, Calgary ranks second in terms of correlation with cities in the United States, which mainly stems from the fact that Canada and the United States are geographically close and share similarities in culture and city structure. On the other hand, we can see that the cities on different continents (North America and Asia) do not share much in common. In particular, we can see Calgary and Singapore exhibit low correlation, and the underlying reason is that Asian cities and North American cities have different urban structures

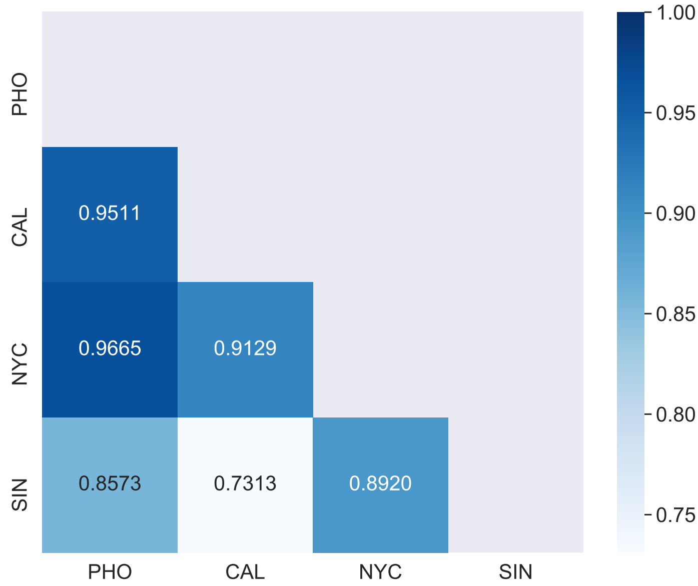


Figure 3.2: The correlation of POIs at category level between four cities, where PHO, CAL, NYC and SIN are short for Phoenix, Calgary, New York and Singpore, respectively.

due to the different cultures of their inhabitants. It can also be noted from Fig. 3.1 that Singapore has a much higher number of POIs in both Residence type and College & University type than the other three cities, whereas the number of POIs in Nightlife spot is less. This makes a clear difference between the correlation of Singapore and the other three cities regarding POIs at category level.

However, the correlation of POIs at category level between cities does not indicate that users behave in the same way. Our main goal is to predict the next POIs that users may visit by referring to their historical check-in records. Therefore, it is the pattern of user check-in across the four cities that can better reflect and unveil user behavior. Based on this, we further investigate the correlation of behavioral patterns at category level between cities.

Correlation of Behavioral Patterns at Category Level. We now examine the correlation of user check-in behavior at category-level in different cities. Given two cities, $A = [A_1, A_2 ... A_i]$ and $B = [B_1, B_2 ... B_i]$ denote the vectors of check-ins under all types of category transition, where A_i and B_i are the number of check-ins under category transition pattern $c_i \rightarrow c_j$ in city A and city B, respectively. Based on

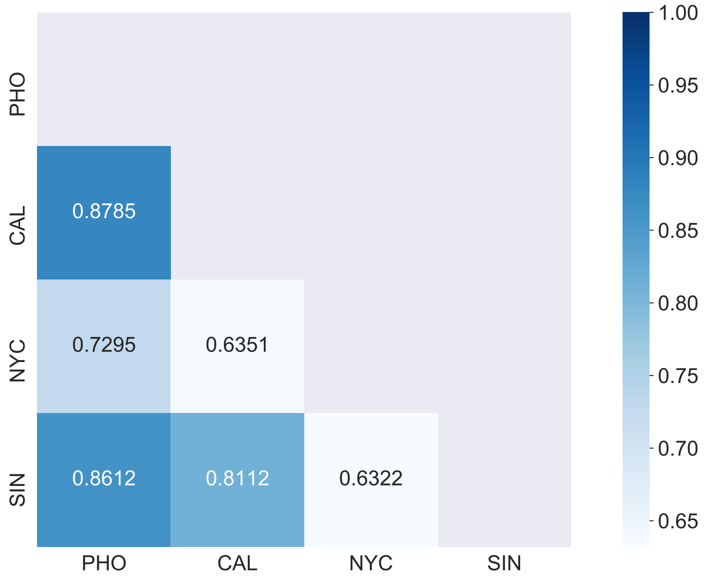


FIGURE 3.3: The correlation of behavioral patterns at category level between four cities.

Eq.(3.1), we calculate the correlation of behavior patterns at category level, and the results are depicted in Fig. 3.3.

Interestingly, the correlation between the four cities regarding behavioral patterns is quite different from that w.r.t. POIs. Specifically, the correlation between Calgary and Phoenix is quite high. On the contrary, the correlation between Singapore and New York is the lowest. To make it clearer how the four cities are correlated and different regarding behavioral patterns at category level, we step further to compare the two most (i.e., Calgary and Phoenix) and least (i.e., New York and Singapore) correlated cities separately. For ease of presentation, we have selected 10 most frequent category transition patterns for comparison as shown in Fig. 3.4 and Fig. 3.5. The x-axis shows the 10 most frequent category transition patterns, e.g., AE2CU (i.e., Arts & Entertainment → College & University); and the $y-$axis shows the proportion of such a transition within a specific city.

Fig. 3.4 compares the proportion of different category transition patterns in the two most correlated cities, i.e., Calgary and Phoenix. Particularly, we can see that users in Calgary and Phoenix tend to go to Nightlife spot after coming out of Shop & Services.

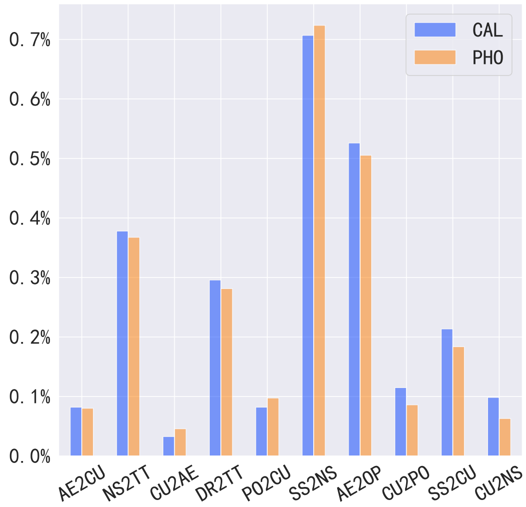


Figure 3.4: Two most correlated cities.

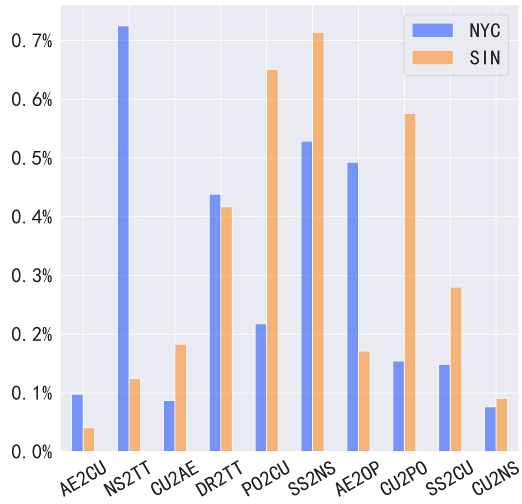


Figure 3.5: Two least correlated cities.

Besides, users in both cities tend to go to Professional Places after going to Arts & Entertainment.

Fig. 3.5 compares the proportion of different category transition patterns in the two least correlated cities, i.e., New York and Singapore. Specifically, users in New York are more likely to take transportation after visiting Nightlife spot, the frequency of which is four times larger than Singapore users. Meanwhile, a large proportion of users in Singapore go to College & University after visiting Professional Places, while the proportion in New York is quite small. Furthermore, a much higher proportion of Singaporean users travelled from College & University to Professional Places in comparison with New York users.

Based on the above observations, we can see that the four cities exhibit correlations to some extent in terms of both POI and behavioral patterns at category level. This, therefore, inspires us to leverage data-rich cities to facilitate data-sparse cities, so as to further improve the performance of the next POI recommendation. However, there are also dissimilarities between cities due to the inherent diversity of culture, urban structure and geographical location. Consequently, instead of blindly transferring all check-ins in data-rich cities to augment data-sparse cities, we hold the principle of "paying more attention to more correlated knowledge". Accordingly, a delicately designed meta-learning framework named MERec has been delivered for performance-enhanced next POI recommendation, which will be introduced in the next chapter.

It is worth noting that since we are aiming at predicting where users are likely to go next based on their historical check-in records, the correlation of behavioral patterns at category level is more in line with our investigated question. Therefore, we adopt such correlation to guide our model design hereafter.

4

The MERec Framework

In this chapter, we first provide preliminaries and definitions for our investigated research question, i.e., meta-learning enhanced next POI recommendation. Then, we introduce the overall framework of our proposed meta-learning framework named MERec. Following that, the detailed components of MERec have been elaborated step by step.

4.1 Preliminaries and Definition

Each city has its unique user set $\mathcal{U}$ and POI set $\mathcal{P}$ without sharing any common users or POIs. For each user u, all his records is ordered by timestamps as in [Zhao et al. (2017)]. Based on this, we then split his historical check-in records as $r = (p, c, g, t)$ into check-in sequences by days, where p is POI ID, c is category ID, g is the GPS location of POI, and t is the timestamp.

- The i-th category sequence of user u is denoted by a set of category tuples, i.e., $C^{u,i} = \{C^u_{t_1}, C^u_{t_2}, ...\}$, where $C^u_{t_k} = (c^u_{t_k}, t^u_k)$.

- The i-th check-in sequence of user u is denoted by a set of POI tuples, i.e., $P^{u,i} = \{P^u_{t_1}, P^u_{t_2}, ...\}$, where $P^u_{t_k} = (p^u_{t_k}, d^u_{t_k}, t^u_k)$, where d_{t_k} is the Euclidean distance between POIs visited at t_{k-1} and t_k.

Given $C^{u,i}$ and $P^{u,i}$, our goal is to predict user u's next POI $p_{t_{k+1}}$ at time t_{k+1}.

Suppose we have a set of *auxiliary cities* $\mathcal{Y}_\mathcal{A} = \{y^{(m)}_{aux} | m \in 1, 2, 3, 4...\}$ and a *target city* $\mathcal{Y}_\mathcal{T} = \{y_{tar}\}$ with a limited amount of check-in sequences, where m is the ID of the corresponding city. Our goal is to transfer knowledge from the auxiliary cities with rich check-in sequences to augment the data-sparse city (i.e., target city), so as to further boost the recommendation performance in the target cities.

In a meta-learning setup, recommendation within each city y_m is regarded as a single task (with its own dataset $\mathcal{D}$). Since a user's previous check-in pattern will have an impact on the present, we divide the check-in records of users in both auxiliary and target cities into a training set and a test set by date (the detailed data splitting process is deferred to Chapter 5). The check-in sequences of $\mathcal{Y}_\mathcal{A}$ and $\mathcal{Y}_\mathcal{T}$ are divided as training sets $\mathbb{D}^{(aux)}_{train}, \mathbb{D}^{(tar)}_{train}$ and test sets $\mathbb{D}^{(aux)}_{test}, \mathbb{D}^{(tar)}_{test}$.

Moreover, each meta-learning tasks has a support set $\mathcal{D}^{spt}$ for training and a query set $\mathcal{D}^{qry}$ for testing. We chronologically select the first several check-in sequences of each user to put into $\mathcal{D}^{spt}$ and the rest into $\mathcal{D}^{qry}$. Finially, our goal is to leverage the training set ($\mathbb{D}^{(aux)}_{train}$) to learn a meta-learner F such that, given the $\mathcal{D}^{spt}$ of a test set, F predicts the parameters θ of recommender f to minimize the recommendation loss $\mathcal{L}$ on the $\mathcal{D}^{qry}$. Formally, it is defined as below:

$$
\begin{aligned}
w^* = arg\,\underset{w}{min} & \sum_{\mathcal{D}=[\mathcal{D}^{spt}, \mathcal{D}^{qry}] \in \mathbb{D}^{(aux)}_{test}} \mathcal{L}(f_\theta, \mathcal{D}^{qry} | \mathbb{D}_{train}, \mathcal{D}^{spt}) \\
& s.t. \quad \theta = F_w(\mathcal{D}^{spt} | \mathbb{D}_{train}),
\end{aligned}
\tag{4.1}
$$

where w, θ are parameters of F and f, respectively; and $\mathbb{D}_{train} = \mathbb{D}^{(aux)}_{train} \cup \mathbb{D}^{(tar)}_{train}$.

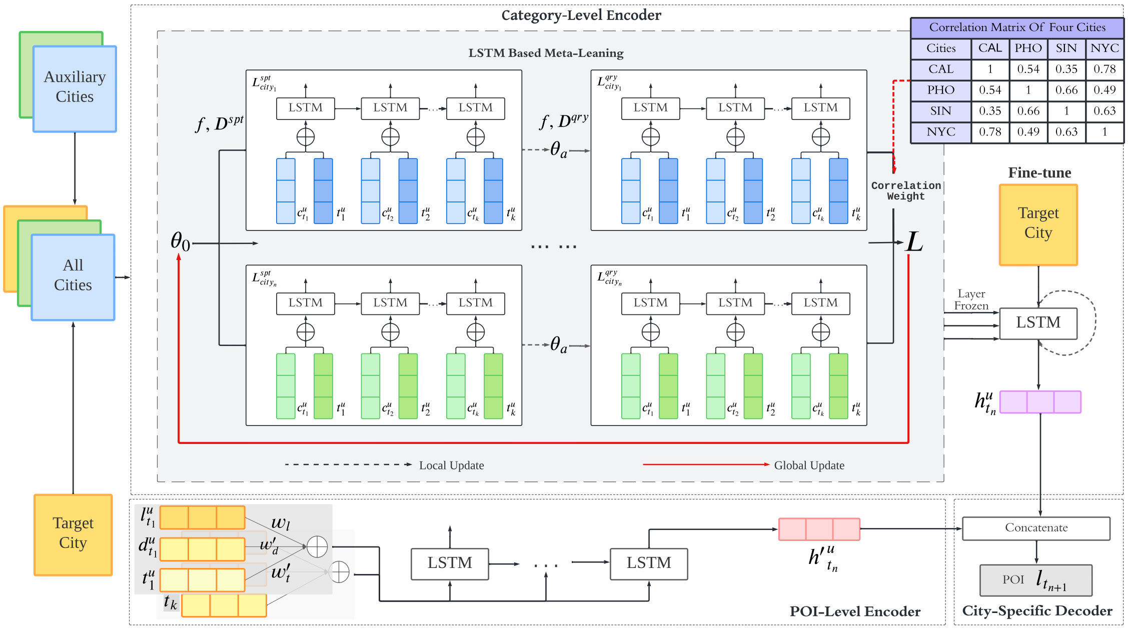


FIGURE 4.1: The overall framework of our proposed MERec.

4.2 The MERec Framework

The overall framework of our proposed MERec is outlined in Fig. 4.1, which is mainly composed of a two-channel encoder (i.e., category-level encoder and POI-level encoder) and a city-specific decoder. In particular, the category-level encoder exploits a meta-learning process to capture common user check-ins transition patterns on category level in each city by holding the principle of "paying more attention to more correlated knowledge". The goal of the POI-level encoder is to learn the accurate POI transition patterns in the target city. Lastly, the city-specific decoder performs the final next POI predictions by concatenating the hidden states of the above two encoders. The overall algorithm of MERec is presented in Chapter 4.2.4.

4.2.1 Base Recommender

We apply the long short-term memory (LSTM) [Hochreiter and Schmidhuber (1997)] as the base recommenders for the two-channel encoder, where $f_\theta^{(cat)}$, $f_\theta^{(poi)}$ denote the base recommenders of category- and POI-level encoders, respectively. The base recommenders are mainly composed of two modules, namely embedding module and output module, as elaborated below.

Embedding Module ($r_i = (p, c, g, t) \longmapsto \mathbf{e}_{hist}$). Regarding the category-level encoder, we simply ignore the user ID and focus on modeling the user historical check-in records at category level. In this module, the embedding matrices E_{cat} and E_{tim} are respectively adopted to map category ID c_i and the timestamp t_i (divide into 24 hours for a day) into latent representations with dimension beging d, which are then concatenated to form the embedding vector $\mathbf{e}^{cat}$ of each record, where $\mathbf{e}^{cat}_{hist}$ is the embeddings of the historical check-in sequence at category level. Similarly, for the POI-level encoder, the embedding matrices $E_{user}, E_{poi}, E_{tim}, E_{dis}$ are adopted to map user ID u_i, POI ID p_i, the timestamp t_i and Euclidean distance d_i (the distance between two consecutive POIs) into d-dimensional vectors, respectively, which are then concatenated to form the embedding vector $\mathbf{e}^{(poi)}$ of each record, where $\mathbf{e}^{(poi)}_{hist}$ is the embeddings of the historical check-in sequence at POI level.

Output Module ($\mathbf{e}_{hist} \longmapsto \hat{y}_i$). The input embeddings are then fed into the LSTM to predict the probability distribution $\hat{y}_i$ (either $\hat{y}_i^{(poi)}$ or $\hat{y}_i^{(cat)}$) on all categories or POIs in the city, denoted as,

$$\hat{y}_i = LSTM(\mathbf{e}_{hist}) \tag{4.2}$$

Note that, for category-level encoder, the parameters of $f_\theta^{(cat)}$ is later meta-learned as introduced in the next subchapter.

4.2.2 Cateogry-Level Encoder

In this step, our goal is to train a sequential model to predict the possible next category by capturing the historical category transition patterns. We mainly utilize LSTM and extend model-agnostic meta-learning (MAML) as the framework for meta-learning update [Finn et al. (2017)] in our scenario.

Meta-learning Setup. Intuitively, MAML learns w initialized by θ_0 of the base recommender $f_\theta^{(cat)}$, which could adapt to new tasks by few update steps on few support samples, and predict well on the query samples. To be specific, each iteration of MAML include two phases: local update and global update on a sampled task batch, where the first phase updates θ_0 locally on the $\mathcal{D}^{spt}$ of each task, and the second phase globally updates θ_0 by gradient descent to minimize the sum of loss on the $\mathcal{D}^{qry}$ of all tasks.

- **Local Update.** Firstly, we sample a batch of cites from $\mathbb{D}_{train}$ where each city y_m has its unique user set $\mathcal{U}_{y_m}$ and POI set $\mathcal{V}_{y_m}$. Then we randomly sample a group of users and from $\mathcal{D}^{spt}_{y_m}$ and $\mathcal{D}^{qry}_{y_m}$. Next, we calculate the training loss on $\mathcal{D}^{spt}_{y_m}$ and locally updated $\theta^{(cat)}$ by one step:

$$\theta'_{y_m} = \theta - \alpha \nabla_\theta \mathcal{L}_{y_m}(f^{(cat)}_\theta, \mathcal{D}^{spt}_{y_m}), \tag{4.3}$$

  where $\mathcal{L}$ is the cross-entropy loss; α is the local learning rate; and $\theta\prime_y$ is the locally updated recommender parameters on each city. For ease of illustration, we omit the superscript (cat) for θ in Eqs. (4.3-4.4).

- **Global Update.** In the second phase, we start with calculating the testing loss on each $\mathcal{D}^{qry}_{y_m}$ with the corresponding θ'_{y_m}. The normal MAML global updating aims to update the initialization θ by one gradient step on the sum of all the testing losses, defined as,

$$\theta = \theta - \beta \nabla_\theta \sum \mathcal{L}_{y_m}(f_{\theta'_{y_m}}, \mathcal{D}^{qry}_{y_m}), \tag{4.4}$$

  where β is the global learning rate.

Correlation Strategy. As analysed in Chapter 3, there are similar and dissimilar characteristics among different cities. Blindly transferring all check-ins from the data-rich cities to the data-sparse cities may introduce some noise thus hurting the recommendation performance. By holding the principle of "paying more attention to more correlated knowledge", our proposed MERec goes a step further from MAML by taking into account the *Correlation of Behavioral Patterns at Category Level* in different cities in the global update. To be specific, we obtain a city correlation matrix γ_{cor} based on the behavioural patterns of the users. In the meta-learning global update, we attentively adapt the gradient across cities based on their correlation. In other words, if the auxiliary city is more correlated to the target city, we adapt the gradient so that it updates faster in that direction. Therefore, the Eq.(4.3) is reformulated as :

$$\theta'_{ym} = \theta - \alpha \nabla_\theta [\mathcal{L}_{ym}(f_\theta^{(cat)}, \mathcal{D}_{ym}^{spt}) \times \gamma_{cor}]. \tag{4.5}$$

By doing so, a more transferable initialization of θ for fast adaption to target cities can be learned after adequate meta-learning iterations. The meta-optimization cross cities are calculated using check-in sequences on each cities with a meta-training step size β. This updates the parameters of the original model so that a few gradient steps are sufficient to tune the parameters to a specific target city.

Freezing Layers and Model Fine Tuning. After obtaining the general representation of auxiliary cities by the meta-learning process, we construct a new model and fine tune with the check-ins of the target city only. Assuming the LSTM model contains L layers, our objective with freezing layers is to maximize the reusability of the general parameters. For this purpose, both Bonab et al. (2021) and Finn et al. (2017) studied the similarity of layers between an adapted model and the general model, whereby they suggested that the main body of the network barely changes and all the adaptation happens in the head layers of the network. Inspired by this finding, we freeze the first l layers of the LSTM network ($1 \leqslant l \leqslant L$), while adding n layers after the l layers and fine tune with target city data only. Hence, $\mathbf{e}_{hist}^{(tar-cat)}$ is fed into the recurrent layer to infer the hidden state $h_{t_k}^u$ of category at t_k, given by,

$$h_{t_k}^u = LSTM_{frozen}(\mathbf{e}_{hist}^{(tar-cat)}). \tag{4.6}$$

We believe that our freezing-layer operation helps generate a network, which is capable of better balancing the parameters between the auxiliary cities and target city after the final model fine-tuning. Meanwhile, the impact of the number of freezing layers l is experimentally explored in our experiments as shown in Chapter 5.4. The freezing layers operation for each target city is shown in line 13 of Algorithm 1.

4.2.3 POI-level Encoder

In order to adapt to the target city, another channel is to train a POI-level model by incorporating the check-in data at POI level of the target city. In this channel, we train the basic LSTM and adopt the check-in sequences at POI-level in the target city as input. In particular, in order to be able to obtain more accurate representation of POIs for the target city, we concatenate the embeddings of user ID, POI ID, timestamp, and distance between two consecutive visited POIs of each record as input. Accordingly, the final embedding of historical check-in sequence at POI level, i.e., $\mathbf{e}_{hist}^{(tar-poi)}$, is then fed into the LSTM to infer the hidden state $\tilde{h}_{t_k}^u$ of POI at t_k, given by,

$$\tilde{h}_{t_k}^u = LSTM_{poi}(\mathbf{e}_{hist}^{(tar-poi)}) \tag{4.7}$$

4.2.4 City-specific Decoder

The city-specific decoder aims to perform next POI prediction based on the latent representations earned from the two-channel encoder. Given a check-in record r_k, based on the cross entropy loss, the objective function $\mathcal{J}$ of next POI prediction task is defined by:

$$\hat{y}_{poi} = softmax(f(h_{t_k}^u; \tilde{h}_{t_k}^u)) \tag{4.8}$$

$$\mathcal{J}_{r_k} = -\sum_{i=1}^{|\mathcal{V}^{(tar)}|} \mathbf{p}_k[i] \cdot log(\hat{y}_{poi}[i]) \tag{4.9}$$

where f is a fully connected layer to transform $(h_{t_k}^u; \tilde{h}_{t_k}^u)$ into a $|\mathcal{V}^{(tar)}|$-dimensional vector; $|\mathcal{V}^{(tar)}|$ is the total number of POIs in the target city; $\hat{y}_{poi}$ represents the predicted probability distribution on all POIs in target city; $\mathbf{p}_k$ is an one-hot embedding of the ground-truth POI p_k. Algorithm 1 summarizes the training learning process of MERec, which is mainly composed of three parts: meta training (lines 3-11), freezing layers and model fine tuning (lines 12-14), as well as next POI prediction (lines 15-16).

Algorithm 1 Mete-learning Enhanced Next POI Recommendation (MERec)

Require: $\mathbb{D}_{train}^{(aux)}$, $\mathbb{D}_{train}^{(tar)}$; base recommender $f_\theta^{(cat)}$ and $f_\theta^{(poi)}$; learning rates α, β; number of shots N; max step of iterations M;

Randomly initialize parameters $\theta = \theta^{(cat)} \cup \theta^{(poi)}$;

Calculate the correlation of behavioral patterns at category level by Eq.(3.1);

while not done **do**

  for *all* $\mathbb{D}_i \in \mathbb{D}_{train}^{(aux)} \cup \mathbb{D}_{train}^{(tar)}$ **do**

    Sample N historical check-ins from $\mathbb{D}_i$ as the adapt_batch;

    Evaluate: $\nabla_\theta \mathcal{L}_{y_m}(f_\theta^{(cat)}, \mathcal{D}_{y_m}^{spt})$ using adapt_batch;

    Calculate the gradient update of θ'_{y_m} by Eq.(4.5);

    Sample another N historical check-ins from $\mathbb{D}_i$ as the eval_batch;

  end for

  Update θ using eval_batch by Eq.(4.4);

end while

Freeze the first l layers as new LSTM model, i.e., LSTM_{frozen};

Fine-tune LSTM_{frozen} using only the category-level training data $\mathbb{D}_{train}^{(tar-cat)}$;

Get the hidden state of POI-level encoder shown in Eq. (4.7) using $\mathbb{D}_{train}^{(tar-poi)}$;

Predict next possible POI via Eq.(4.8);

Calculate the prediction loss for each check-in record via Eq.(4.9);

5

Experiments and Results

In this chapter, we conduct the experiments on four real-world datasets from Foursquare as introduced in Chapter 3 to evaluate the performance of our proposed MERec on the next POI recommendation task[1]. Our experimental evaluation is designed to answer three research questions (**RQs**).

- **RQ1**: Does MERec outperform other state-of-the-art methods for the next POI recommendation?
- **RQ2**: How do different components of MERec affect its performance?
- **RQ3**: How do hyper-parameter settings affect MERec?

In what follows, we introduce the experimental settings, and present the experimental results followed by the corresponding in-depth analysis.

[1]Our code is released at https://github.com/OliverWang-Au/DAMER_Framework.

5.1 Experimental Setup

Datasets. We conduct extensive experiments on the four datasets as shown in Table 1.1. We use one of the cities as the target city and the other cities as auxiliary cities in the experiment. Following [Huang et al. (2019)], we treat the first 80% sequences of each user as training set, the latter 10% as the validation set and the last 10% as test set. Note that we filtered POIs with less than three check-ins and users with less than five interactions, respectively.

Evaluation Metrics. We adopt two widely-used ranking evaluation metrics: *Hit Ratio* at K ($HR@K$) and *Normalized Discounted Cumulative Gain* at K ($NDCG@K$). $HR@K$ measures whether the test POI shows within the top-K ranked list while the NDCG@K takes the position of the test POI into account and penalizes the score if it is ranked lower in the list.

Comparison Baselines. In order to verify the effectiveness of our proposed method, we compare the following seven state-of-the-art next POI recommendation approaches, which can be divided into three groups: traditional methods (TM), deep learning methods (DL), and meta-learning based methods (META).

Specifically, we consider two traditional methods.

- **MostPop**: recommends next POI based on the popularity of POIs.
- **BPRMF** [Rendle et al. (2012); Yuan et al. (2016)]: matrix factorization based method, optimized via Bayesian personalized ranking.

Three classic deep learning methods are also compared.

- **NeuMF** [He et al. (2017)]: an item recommendation model combining matrix factorization with MLP.
- **ATST-LSTM** [Huang et al. (2019)]: a recent next POI recommender, attending user embedding on the LSTM outputs with distance and delta time between successive check-ins considered.

- **iMTL** [Zhang et al. (2021)]: a recently-proposed approach for next POI recommendation, using two-channel encoder and a task-specific decoder for capturing the sequential correlations of activities and location preferences.

Meanwhile, three meta-learning based methods have been taken into account.

- **MAML** [Finn et al. (2017)]: a model-agnostic framework based meta-learning for few shot learning.
- **CHAML** [Chen et al. (2021)]: a recent framework incorporating hard sample mining and curriculum learning into meta-learning step.

Hyper-parameter Settings. The optimal hyper-parameter settings for all methods are empirically found out based on the performance on the validation set. In particular, the embedding size is searched from $\{32, 64, 128, 256\}$; for BPRMF and deep learning based methods, we set the batch size as 256; the learning rate is selected from { 0.1, 0.05, 0.01, 0.005, 0.001, 0.0001. Accordingly, for all META methods, the learning rates α and β are selected from {0.5, 0.1, 0.01, 0.001, 0.0001}; and the batch size is set as 256 to ensure fair comparison. For our freezing step, we vary the number of freezing layers in the range of $[1, 4]$ stepped by one. Finally, we select freezing 3 layers of LSTM for all datasets.

5.2 Performance Comparison (RQ1)

The comparative results of different methods are presented in Table 5.1, where the best results are highlighted in bold; the runner-up is underlined; and the column 'Improve' indicates the improvements achieved by our proposed MERec relative to the runner up. Next, we analyze the results aiming to answer the first research question, i.e., **RQ1**.

In terms of the four datasets, the traditional methods (MostPop, BPRMF) generally perform worse than deep learning methods (NeuMF, ATST-LSTM, iMTL) demonstrating the efficacy of neural network on more accurate recommendation. As for RNN based methods, ATST-LSTM and iMTL performs better than NeuMF, which indicates the

TABLE 5.1: Comparative results of all approaches on the four datasets, where 'H' refers to 'Hit Ratio' and 'N' means 'NDCG'; the best results are highlighted in bold and the runner up is underlined; the column 'Improve' indicates the improvements achieved by MERec relative to the runner up.

		TM		DL			META			*Improve*
		MostPop	BPRMF	NeuMF	ASTA-LSTM	iMTL	MAML	CHAML	MERec	
Calgary	H@5	0.0988	0.1304	0.1431	0.2924	0.2652	0.3987	<u>0.3995</u>	**0.4274**	6.98%
	H@10	0.1547	0.2349	0.2368	0.3705	0.3184	0.4618	<u>0.4777</u>	**0.5054**	5.80%
	N@5	0.0632	0.0928	0.0989	0.2134	0.1857	<u>0.3178</u>	0.3093	**0.3378**	6.29%
	N@10	0.0814	0.1672	0.1669	0.2383	0.2299	<u>0.3362</u>	0.3315	**0.3564**	6.01%
Phoenix	H@5	0.0682	0.1093	0.1316	0.2366	0.2410	0.3549	<u>0.3660</u>	**0.3928**	7.32%
	H@10	0.1068	0.1584	0.1852	0.3125	0.3370	<u>0.4508</u>	0.4419	**0.4531**	0.51%
	N@5	0.0419	0.0688	0.0869	0.1635	0.1753	0.2633	<u>0.2648</u>	**0.2796**	5.59%
	N@10	0.0547	0.0848	0.1042	0.1883	0.2065	<u>0.2949</u>	0.2891	**0.2993**	1.49%
Singapore	H@5	0.0365	0.0848	0.1004	0.2165	0.2388	0.2991	<u>0.3571</u>	**0.3705**	3.74%
	H@10	0.0635	0.1450	0.1696	0.2879	0.3080	0.3816	<u>0.4486</u>	**0.4488**	0.04%
	N@5	0.0231	0.0452	0.0697	0.1532	0.1696	0.2188	<u>0.2650</u>	**0.2707**	2.15%
	N@10	0.0318	0.0648	0.0925	0.1760	0.1922	0.2451	<u>0.2981</u>	**0.2998**	0.57%
New York	H@5	0.0214	0.0558	0.0959	0.1763	0.2187	0.2456	<u>0.2745</u>	**0.2991**	8.96%
	H@10	0.0336	0.0994	0.1495	0.2455	0.2879	0.3373	<u>0.3526</u>	**0.3995**	13.3%
	N@5	0.0134	0.0265	0.0595	0.1257	0.1484	0.1652	<u>0.1865</u>	**0.2107**	12.98%
	N@10	0.0173	0.0237	0.0770	0.1485	0.1705	0.2072	<u>0.2118</u>	**0.2436**	15.01%

capability of RNN on modeling the sequential dependency. iMTL performs better than ATST-LSTM, as it leverages multi-task learning (MTL) framework to jointly learn user preference on both activities (i.e., categories) and POIs, which exhibits the superiority of MTL on better next POI recommendation. Unsurprisingly, meta-learning based methods (MAML, CHAML) bring further enhancement compared with other methods, owing to the specialized and efficient design of transferring knowledge for alleviating the data sparsity issue.

Overall, our proposed model MERec observably outperforms all the other baselines, including both deep learning and meta-learning recommenders. Specifically, the relative improvements over the runner-up baseline is respectively 6%, 4%, 2% and 4% on four cities across the four metrics on average. This helps further confirms the benefits

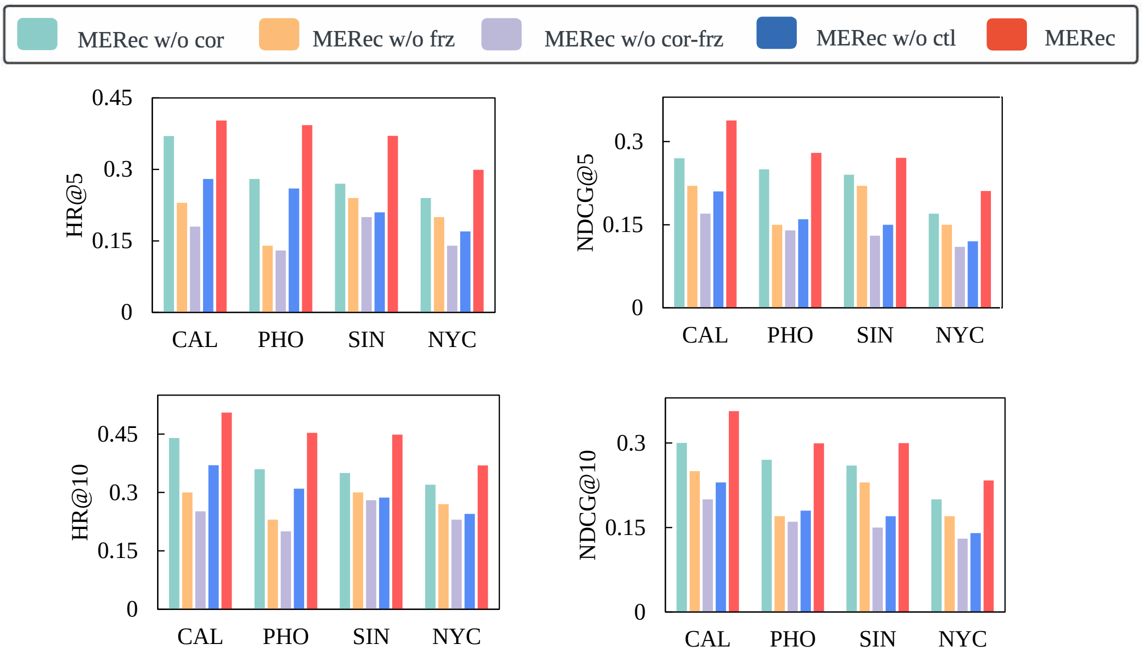


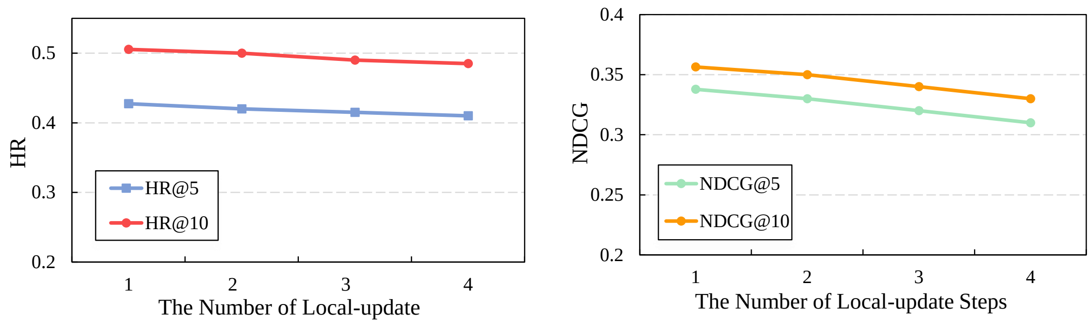
Figure 5.1: Performance comparison for variants of MERec on the four datasets.

of (1) leveraging check-ins of auxiliary cities to augment target cities, and (2) paying more attention to more correlated knowledge, when training the meta-learner.

5.3 Ablation Study (RQ2)

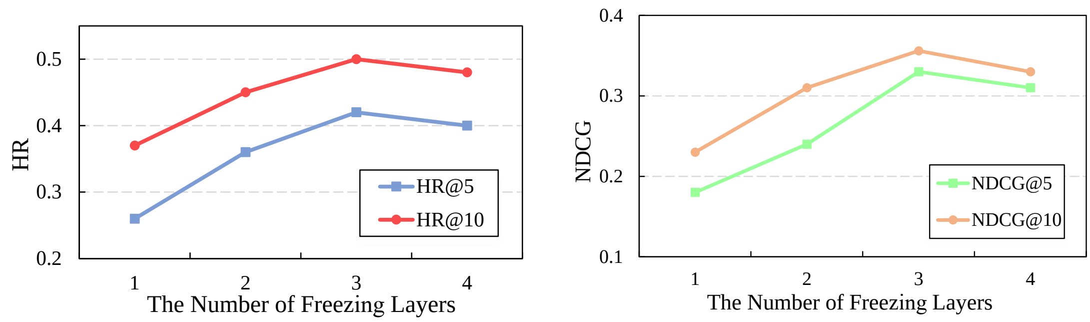
To answer **RQ2**, we analyze the contributions of different components in MERec by comparing following variants:

- $MERec_{w/o\ cor}$: removes the correlation strategy from the meta-learner.
- $MERec_{w/o\ frz}$ removes the freeze and fine-tune operation from the category-level encoder.
- $MERec_{w/o\ cor-frz}$ removes both correlation strategy and freeze/fine-tune operation from the meta-learner and category-level encoder, respectively.
- $MERec_{w/o\ ctl}$ removes the category-level encoder, but only retains the POI-level encoder.



(a) (b)

FIGURE 5.2: The impact of local-update steps of Meta-learning.



(a) (b)

FIGURE 5.3: The impact of the number of freezing layers.

We report the results of different variants in Fig. 5.1, where MERec significantly outperforms its variants regarding various metrics across the four datasets. In particular, we notice that MERec$_{w/o\ cor-frz}$ performs worse than either MERec$_{w/o\ cor}$ and MERec$_{w/o\ frz}$, which suggests that both the correlation strategy and freeze/fine-tune operation indeed improve the recommendation performance. Generally, the performance decrease of MERec$_{w/o\ frz}$ far exceeds that of MERec$_{w/o\ cor}$, implying that the freeze and fine-tune operation plays a more important role than the correlation strategy. Besides, it is worth noting that MERec$_{w/o\ ctl}$ underperforms MERec, which further helps confirms the advantages of meta-learning with auxiliary check-ins and correlation-aware strategy. To sum up, our proposed MERec benefits from the three delicately designed components.

5.4 Parameter Sensitivity Analysis (RQ3)

To answer **RQ3**, we investigate the influence of different hyper-parameters. Particularly, we analyze the impacts of two key parameters of MERec, i.e., the number of local-update steps in Eq.(4.3) and the number of freezing layers in Chapter 4.2.2. For illustration, we only show the results on Calgary dataset, and similar trends can be observed on the rest three datasets.

Figs. 5.2 (a-b) depict the model performance w.r.t. the number of local-update steps. We empirically find out that updating only one step is sufficient to obtain better recommendation accuracy, which also increases the model efficiency. Figs. 5.3 (a-b) display the influence of the number of layers frozen on the model performance. We vary the number of layers frozen in the range of $[1, 4]$ stepped by one. As observed, with the layer increasing, the performance first goes up and then drops slightly. The best setting for the number of freezing layer is 3 on the four datasets.

6
Conclusion

In this paper, we propose a Meta-learning Recommendation (MERec) framework for the next POI recommendation by leveraging check-ins from auxiliary cities to augment the target cities, and holing the principle of "paying more attention to more correlated knowledge". In particular, we devise a two channel encoder, i.e., category-level encoder and POI-level encoder, to capture the transition patterns of categories and POIs, whereby a city-correlation based strategy is devised to attentively capture common knowledge (i.e., patterns) from auxiliary cities. The city-specific decoder then concatenates the latent representations of the two-channel encoder to perform next POI prediction for the target city. Extensive experiments on four real-word datasets across different evaluation metrics demonstrate the superiority of our proposed MERec. In addition, due to the limitation of computing power, we only conducted experiments on the dataset of four cities. Also based on literature review, we only chose LSTM as the base recommender. Therefore, in future research we plan to extend MERec on three directions: 1) We plan to add more different city datasets to experiment 2) we plan to expand other sequential based model to boost the improvement. 3) We plan to consider the use of correlation strategy as self-learning parameter for learning.

References

R. P. Adams, G. E. Dahl, and I. Murray. Incorporating side information in probabilistic matrix factorization with gaussian processes. *arXiv preprint arXiv:1003.4944*, 2010. 5

H. Bonab, M. Aliannejadi, A. Vardasbi, E. Kanoulas, and J. Allan. Cross-market product recommendation. *arXiv preprint arXiv:2109.05929*, 2021. 3, 24

Y. Chen, X. Wang, M. Fan, J. Huang, S. Yang, and W. Zhu. Curriculum meta-learning for next poi recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2692–2702, 2021. 2, 3, 8, 29

C. Cheng, H. Yang, I. King, and M. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 17–23, 2012. 6

C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *23rd International Joint Conference on Artificial Intelligence*, 2013. 2, 6

Y. Cui, H. Sub, Y. Zhao, H. Yin, and K. Zheng. Sequential-knowledge-aware next poi recommendation: A meta-learning approach. 2021. 3, 8

J. Ding, G. Yu, Y. Li, D. Jin, and H. Gao. Learning from hometown and current city: Cross-city poi recommendation via interest drift and transfer learning. *Proceedings of* *the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–28, 2019. 3

A. Farseev, I. Samborskii, A. Filchenkov, and T.-S. Chua. Cross-domain recommendation via clustering on multi-layer graphs. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 195–204, 2017. 7

S. Feng, X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan. Personalized ranking metric embedding for next new poi recommendation. In *24th International Joint Conference on Artificial Intelligence*, 2015. 6

C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. 22, 24, 29

H. Gao, J. Tang, X. Hu, and H. Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 93–100, 2013. 6

Q. Gu, J. Zhou, and C. Ding. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 199–210. SIAM, 2010. 5

X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 173–182, 2017. 28

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9 (8):1735–1780, 1997. 7, 21

L. Huang, Y. Ma, S. Wang, and Y. Liu. An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Transactions on Services Computing*, 2019. 2, 28

M. Islam, M. M. Mohammad, S. S. S. Das, M. E. Ali, et al. A survey on deep learning based point-of-interest (poi) recommendations. *arXiv preprint arXiv:2011.10187*, 2020. 1

T. Kurashima, T. Iwata, T. Hoshide, N. Takaya, and K. Fujimura. Geo topic model: joint modeling of user's activity area and interests for location recommendation. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pages 375–384, 2013. 5

J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *2012 IEEE 28th International Conference on Data Engineering*, pages 450–461. IEEE, 2012. 5

D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 831–840, 2014. 2, 6

B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pages 1043–1051, 2013. 6

Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *30th AAAI Conference on Artificial Intelligence*, 2016a. 7

Y. Liu, C. Liu, B. Liu, M. Qu, and H. Xiong. Unified point-of-interest recommendation with temporal interval assessment. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1015–1024, 2016b. 6

Y. Liu, A. Pei, F. Wang, Y. Yang, X. Zhang, H. Wang, H. Dai, L. Qi, and R. Ma. An attention-based category-aware gru model for the next poi recommendation. *International Journal of Intelligent Systems*, 36(7):3174–3189, 2021. 2

T. Man, H. Shen, X. Jin, and X. Cheng. Cross-domain recommendation: An embedding and mapping approach. In *IJCAI*, volume 17, pages 2464–2470, 2017. 7

N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel. A simple neural attentive meta-learner. *arXiv preprint arXiv:1707.03141*, 2017. 8

T. Munkhdalai, X. Yuan, S. Mehri, and A. Trischler. Rapid adaptation with conditionally shifted neurons. In *International Conference on Machine Learning*, pages 3664–3673. PMLR, 2018. 8

E. Palumbo, G. Rizzo, R. Troncy, and E. Baralis. Predicting your next stop-over from location-based social network data with recurrent neural networks. In *RecTour@ RecSys*, pages 1–8, 2017. 7

T. Qian, B. Liu, Q. V. H. Nguyen, and H. Yin. Spatiotemporal representation learning for translation-based poi recommendation. *ACM Transactions on Information Systems (TOIS)*, 37(2):1–24, 2019. 2

S. Ravi and H. Larochelle. Optimization as a model for few-shot learning. 2016. 8

S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012. 28

J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30, 2017. 8

Z. Sun, C. Li, Y. Lei, L. Zhang, J. Zhang, and S. Liang. Point-of-interest recommendation for users-businesses with uncertain check-ins. *IEEE Transactions on Knowledge and Data Engineering*, 2021. 13

H. Tan, D. Yao, T. Huang, B. Wang, Q. Jing, and J. Bi. Meta-learning enhanced neural ode for citywide next poi recommendation. In *2021 22nd IEEE International Conference on Mobile Data Management (MDM)*, pages 89–98, 2021. 2, 3, 8

J. Vanschoren. Meta-learning: A survey. *arXiv preprint arXiv:1810.03548*, 2018. 8

J. Wang, Y. Ren, J. Li, and K. Deng. The footprint of factorization models and their applications in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 40(4):1–32, 2021. 2

X. Wang, Z. Peng, S. Wang, S. Y. Philip, W. Fu, and X. Hong. Cross-domain recommendation for cold-start users via neighborhood based feature mapping. In *International Conference on Database Systems for Advanced Applications*, pages 158–165. Springer, 2018. 7

M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 15–24, 2016. 2

D. Yang, D. Zhang, and B. Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3):1–23, 2016. 11

H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In *The World Wide Web Conference*, pages 2181–2191, 2019. 8

J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 171–179. SIAM, 2013. 7

M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 458–461, 2010. 6

M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and Development in Information Retrieval*, pages 325–334, 2011. 6

F. Yuan, G. Guo, J. M. Jose, L. Chen, H. Yu, and W. Zhang. Lambdafm: learning optimal ranking with factorization machines using lambda surrogates. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 227–236, 2016. 28

Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 363–372, 2013. 6

H. Zhang, S. Wei, X. Hu, Y. Li, and J. Xu. On accurate poi recommendation via transfer learning. *Distributed and Parallel Databases*, 38(3):585–599, 2020a. 3

J.-D. Zhang and C.-Y. Chow. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 334–343, 2013. 6

J.-D. Zhang, C.-Y. Chow, and Y. Li. Lore: Exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 103–112, 2014. 6

L. Zhang, Z. Sun, J. Zhang, H. Kloeden, and F. Klanner. Modeling hierarchical category transition for next poi recommendation with uncertain check-ins. *Information Sciences*, 515:169–190, 2020b. 11, 12, 13

L. Zhang, Z. Sun, J. Zhang, Y. Lei, C. Li, Z. Wu, H. Kloeden, and F. Klanner. An interactive multi-task learning framework for next poi recommendation with uncertain check-ins. In *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*, pages 3551–3557, 2021. 2, 7, 11, 12, 29

S. Zhao, T. Zhao, H. Yang, M. R. Lyu, and I. King. Stellar: Spatial-temporal latent ranking for successive point-of-interest recommendation. In *30th AAAI Conference on Artificial Intelligence*, 2016. 6

S. Zhao, T. Zhao, I. King, and M. R. Lyu. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 153–162, 2017. 19