Topic-Enhanced Memory Networks for Personalised Point-of-Interest Recommendation

Xiao Zhou
Computer Laboratory
University of Cambridge
Cambridge, UK
xz331@cam.ac.uk

Cecilia Mascolo
Computer Laboratory
University of Cambridge
Cambridge, UK
cm542@cam.ac.uk

Zhongxiang Zhao
WeChat Business Group
Tencent Inc.
Beijing, China
abnerzxzhao@tencent.com

ABSTRACT

Point-of-Interest (POI) recommender systems play a vital role in people's lives by recommending unexplored POIs to users and have drawn extensive attention from both academia and industry. Despite their value, however, they still suffer from the challenges of capturing complicated user preferences and fine-grained user-POI relationship for spatio-temporal sensitive POI recommendation. Existing recommendation algorithms, including both shallow and deep approaches, usually embed the visiting records of a user into a single latent vector to model user preferences: this has limited power of representation and interpretability. In this paper, we propose a novel topic-enhanced memory network (TEMN), a deep architecture to integrate the topic model and memory network capitalising on the strengths of both the global structure of latent patterns and local neighbourhood-based features in a nonlinear fashion. We further incorporate a geographical module to exploit user-specific spatial preference and POI-specific spatial influence to enhance recommendations. The proposed unified hybrid model is widely applicable to various POI recommendation scenarios. Extensive experiments on real-world WeChat datasets demonstrate its effectiveness (improvement ratio of 3.25% and 29.95% for context-aware and sequential recommendation, respectively). Also, qualitative analysis of the attention weights and topic modeling provides insight into the model's recommendation process and results.

CCS CONCEPTS

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

KEYWORDS

Recommender Systems; Neural Networks; Topic Modeling

1 INTRODUCTION

In the information explosion era, recommender systems have become increasingly important in our daily life by improving the suggestions we received online based on personal tastes [35]. As a particular type of recommender systems focusing on real-world locations, point-of-interest (POI) recommendation has drawn intensive attention recently with the prevalence of smart mobile devices and the rapid growth of location-based social networks (LBSNs). Specifically, LBSNs, such as Yelp[1] and Foursquare[2] link the physical and virtual worlds by offering users a way to share their life experiences with POIs via the "check-in" function. Through the lens of check-in data in LBSNs, rich knowledge becomes available to mine users' visiting preferences for customised POI recommendation [31]. Such kind of services not only benefit users by providing them appealing venues to explore, but also facilitate targeted advertising with significant economic efficiency enhancement.

Unlike other recommender systems that push digital goods, e.g. e-books, news, and movies, POI recommendation aims at offering users preferred new venues to explore in the physical world [25], which can be largely affected by various real-life factors and thus faces more challenges. Firstly, to experience a POI, a user has to physically visit it, which is generally more costly and time-consuming than watching a movie or listening to a song online. Also, even if the user does visit the venue, she might prefer not to leave a check-in record due to privacy and security concerns [27]. For these reasons, the number of POIs that a user generally interacts with is extremely small compared with the total number of POIs in the LBSN, leading to a very sparse user-POI matrix that plagues a considerable number of current POI recommender systems [29]. Moreover, unlike pure online interactions, users' activities in the physical world are limited by travel distance and time [25], making POI recommendation a more complex issue. Considering the distinguishing characteristics of POI recommendation task, corresponding research efforts have been spend improving its effectiveness. The existing approaches in this area can be roughly grouped into two categories: neighbourhood-based methods and latent factor models [8]. Neighbourhood-based approaches are good at describing local strong associations between users and POIs, but typically ignore the vast majority of user-POI interactions that are not similar enough. While for latent factor models like Matrix Factorization (MF)[15], they can easily extract the global structure of relationships between users and POIs, but have limitations in capturing their local features. Instead of establishing either a neighbourhood-based or latent factor model, we propose a unified hybrid architecture

---

[1] www.yelp.com
[2] https://foursquare.com/

which combines the benefits of both techniques to enrich predictive capabilities by capturing both global and neighbourhood features of users and POIs. Furthermore, aiming to learn higher order complex relations between users and POIs that cannot be obtained through simple functions, deep learning techniques are adopted.

The model proposed in this paper, *Topic-Enhanced Memory Networks* (**TEMN**), is an end-to-end framework for personalised POI recommendation, which consists of three integrated components. Specifically, a memory-augmented neural network is employed to model fine-grained relations between each user-POI pair via a metric learning approach. This memory network (MN) component is linked with a topic modeling module, called temporal latent Dirichlet allocation (TLDA) through user embedding. In this way, TLDA module can not only be used to infer the general preferences of users towards POIs and time slots, but also influence the representations of POIs and users in MN to enhance the recommendation performance. Furthermore, to capture geographical effects in POI recommendation, geographical modeling module is also devised.

Our primary contributions can be summarised as follows:

- We propose an end-to-end deep learning framework that integrates neighbourhood-based and global preferences of users.
- We devise a more flexible architecture that incorporates multiple types of contextual information into POI recommendation and make it applicable to various recommendation scenarios.
- We build a hybrid model which combines supervised and unsupervised learning and capitalises on the advances in both memory networks and topic modeling. Through a mutual learning mechanism, our model is also able to provide users' probability distributions over topics that are influenced by memory network.
- Comprehensive experiments on large WeChat datasets in different recommendation scenarios demonstrate the effectiveness of TEMN against competitive state-of-the-art baselines (improvement ratio of 3.25% and 29.95% for context-aware and sequential recommendation, respectively).
- Besides quantitative improvements, by incorporating neural attention mechanisms and topic model in TEMN, the interpretability of the POI recommendation is significantly promoted.

2 RELATED WORK

To alleviate the sparsity issue of user-POI interaction data and cater for the context awareness nature of POI recommendation, auxiliary information have been incorporated into existing POI recommender systems [25]. For instance, [17] incorporated spatial clustering features of check-ins into a weighted matrix factorization framework. [9, 27, 34] claimed that POIs clustered within the same geographical region should be restricted to similar representations compared with their distant counterparts in recommendation. Another group of researchers focused on how temporal effects can be utilised to improve POI recommendation by considering either temporal cyclic patterns [10, 32] or sequential influence [4, 6, 34]. Besides spatio-temporal information, other types of information have also been explored to facilitate POI recommendation performance, such as social influence [5, 7, 16, 22, 28, 30, 31], categorical information [31, 33], visual content [26], and text information [3, 24] of POIs.

From the angle of recommendation techniques, scholars have proposed some effective models. For example, MF [15] and its

variants have been widely applied to POI recommendation tasks. Vanilla MF was initially used to deal with general POI recommendation, where only user-POI interactions are leveraged. When additional information like spatio-temporal information and social relationship became available, some models derived from basic MF (e.g., IRenMF [19], GeoMF [17], GeoIE [25]) were proposed and adapted to context-aware recommendation. Markov chain models (e.g., LBPR [11], NLPMM [4], FPMC-LR [6], LORE [32]) and recurrent neural network (RNN) (e.g., ST-RNN [18], CARA [20], DRCF [18]) have also been employed when temporal effects are particularly valued in sequential POI recommendation.

3 PROBLEM FORMULATION

In this section, we formulate our research problems and introduce the notations used throughout the paper.

In a POI recommender system, let $\mathcal{U}$ and $\mathcal{V}$ denote the user set and the POI set, respectively. The user-POI interaction matrix $Y = \{y_{uv} \mid u \in \mathcal{U}, v \in \mathcal{V}\}$ is defined according to users' implicit feedback, where each entry y_{uv} in $Y \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}|}$ records whether user u has visited POI v as follows:

$$y_{uv} = \begin{cases} 1, & \text{if interaction } (u, v) \text{ is observed;} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Here a value of 1 for y_{uv} indicates that user u has visited the venue v before. Otherwise, it is assigned a value of 0, meaning that there is no interaction existing between them and the preference of user u for POI v is currently unclear. Mathematically, the main goal of the POI recommender system is to estimate the scores of the unobserved entries in matrix Y for recommendation.

Since our aim is to propose a more generic framework for POI recommendation that is sufficiently flexible to incorporate various contextual information and is widely applicable to different scenarios, we present the definitions of three POI recommendation scenarios involved in this paper as follows:

General POI recommendation. In this scenario, only the interactions between users $\mathcal{U}$ and POIs $\mathcal{V}$ are taken into account. In other words, the interaction matrix Y is taken as the only input to the recommender system. The main task here is to provide each $u \in \mathcal{U}$ a list of POIs consisting of venues that the target user has not explored before, but is potentially interested in.

Sequential POI recommendation. Sequential recommendation predicts successive POIs that are likely to be visited by a user given her check-in history. In such a scenario that sequence matters, for each user u, her past interactions with POIs are firstly sorted according to the check-in timestamps ascendingly. Instead of using the whole visiting sequence of the user directly, we divide it into some shorter segments by setting a time interval threshold ΔT. Specifically, if the time interval between two adjacent check-ins is larger than ΔT, they are not regarded as successive check-ins, since the large time interval may indicate they are irrelevant [32]. Based on this assumption, a check-in sequence set forms for each user and can be used as training data for sequential POI recommendation.

Context-aware POI recommendation. As stated in Section 2, multiple types of rich contextual information associated with check-ins can be used to enhance POI recommender systems. In this paper, we will focus more on spatio-temporal aware POI recommendation.

4 TOPIC-ENHANCED MEMORY NETWORKS

In this section, our proposed model Topic-Enhanced Memory Networks for POI recommendation (TEMN) is introduced. After presenting the general architecture of the model, we will describe its main components respectively before elaborating on how to integrate them with each other for overall training.

4.1 General Framework

In Figure 1, we present a visual depiction of TEMN's overall architecture. It can be seen that at a high level, TEMN consists of three key parts: memory network, TLDA, and geographical modeling. The first two components are linked with each other, which allows for modeling the nonlinear interactions between local features learned from the neighbourhood-based memory network and global preferences extracted from the topic model. With such an architecture, given a user u, the final prediction score for each unvisited POI v is a reflection of both the fine-grained relationship between user-POI pair $<u, v>$ and the general interest of user u. Technically, this capability is realised by the topic-enhanced user embedding, which is made up of two parts: 1) a representation related to the user's memory component that encodes her visiting records (named as memory embedding), and 2) a vector employed to encode her intrinsic preference extracted through topic modeling (named as intrinsic embedding).

4.2 Memory Network

In our framework, the module of memory network is utilised to encode users' previous check-in records, capture the local relationships between venues, and learn the relational vector for each user-POI pair which represents complex interactions between them.

4.2.1 User and POI embedding. In the memory network, POIs are initially represented by binary sparse vectors through one-hot encoding. Then they are further transformed into low-dimensional dense vectors to get POI embeddings. Formally, given a venue v, let $q_v \in \mathbb{R}^d$ denotes its embedding vector. Here d is the dimension size of the POI embedding. Different from POI embedding method, for each user u, her memory embedding $p_u^m \in \mathbb{R}^d$ is derived based on the POIs she has visited V_u^+ as:

$$p_u^m = \frac{1}{|\mathcal{V}_u^+|} \sum q_i^m (i \in \mathcal{V}_u^+) \quad (2)$$

where q_i^m is the embedding vector of POI i. Then for each given user-POI pair $<u, v>$, we calculate their joint embedding e_{uv} through Hadamard product as below:

$$e_{uv} = p_u^m \odot q_v \quad (3)$$

The dimension size of the generated vector e_{uv} is the same as that of p_u^m and q_v. This joint embedding of user-POI pair will be fed into a memory-augmented neural network later to make predictions.

4.2.2 Key-value memory networks. The core of the memory network module consists in a memory matrix $\mathbf{M} \in \mathbb{R}^{h \times d}$, where h represents the number of memory slots and d denotes the size of each memory cell. In our model, this memory matrix stores the information related to user-POI interactions and can be read and updated adaptively by adopting key-value attention mechanisms.

Technically, a key matrix $\mathbf{K} \in \mathbb{R}^{d \times h}$ is employed, where the number of key vectors is h, and the dimensionality of each key slot $k_i \in \mathbf{K}$ is d, which is the same as that in user and POI embedding. Our main aim here is to learn an attention vector w. For each $<u, v>$ pair, when their joint embedding e_{uv} is fed into the networks, the similarity between e_{uv} and each key vector k_i will be first calculated by dot product before being converted to a relevance probability using softmax function:

$$w_i = softmax(e_{uv}^T \cdot k_i) = \frac{exp(e_{uv}^T \cdot k_i)}{\sum_j exp(e_{uv}^T \cdot k_j))}, \forall j = 1, 2, \cdots h \quad (4)$$

where w_i represents the attention weight of the i-th element in the attention vector w, which will be further utilised to collect useful information from each memory slice $m_i \in \mathbf{M}$ to generate the relation vector r_{uv} of the given user-POI pair $<u, v>$ through:

$$r_{uv} = \sum_i w_i m_i, \forall i = 1, 2, \cdots h \quad (5)$$

This obtained latent relation vector r_{uv} is the output of the memory network module.

4.2.3 Optimisation. The latent representation of user-POI interactions plays as a translation vector that enables the model to learn relationships between each specific user-POI pair in metric space adaptively [23]. The scoring function implemented here is based on the idea of TransE [2] model, where we compute the loss of the triple $<p_u^m, r_{uv}, q_v>$ by the Euclidean distance and calculate the score for each user-POI pair $<u, v>$ as:

$$s_{uv}^m = -||p_u^m + r_{uv} - q_v||_2^2 \quad (6)$$

where $||\cdot||_2$ is the $L2$-norm of the vector. s_{uv}^m represents the score of user-POI pair $<u, v>$ predicted through the memory network. The higher the score, the more likely user u would be interested in the recommended venue v.

We optimise the memory network module under the Bayesian personalised ranking (BPR) [21] criterion, where the key assumption is that observed entries should be ranked higher than their unobserved counterparts. Given a positive instance $v \in \mathcal{V}_u^+$ visited by user u, a corrupted example $j \notin \mathcal{V}_u^+$ which the user has not interacted with yet would be selected based on a certain negative sampling rule. Then both of the POIs v and j would go through the same user and POI embedding layer as introduced, respectively. When calculating the score s_{uj}^m for the corrupted pair $<u, j>$, the relation vector used is the same as that generated by positive pair $<u, v>$. The reason for doing like this is that we empirically found it providing much better performance than generating another relation vector for the negative sample. Based on pairwise comparisons, the objective function in this part is defined as:

$$\mathcal{L}^m = \sum_u \sum_{v \in \mathcal{V}_u^+} \sum_{j \notin \mathcal{V}_u^+} max(0, s_{uj}^m - s_{uv}^m + \lambda^m) \quad (7)$$

where λ^m is the margin that separates positive and negative examples. Rectified linear unit (*ReLU*) function is adopted by us as the nonlinear activation function here as it gives better experimental results than using the sigmoid function.
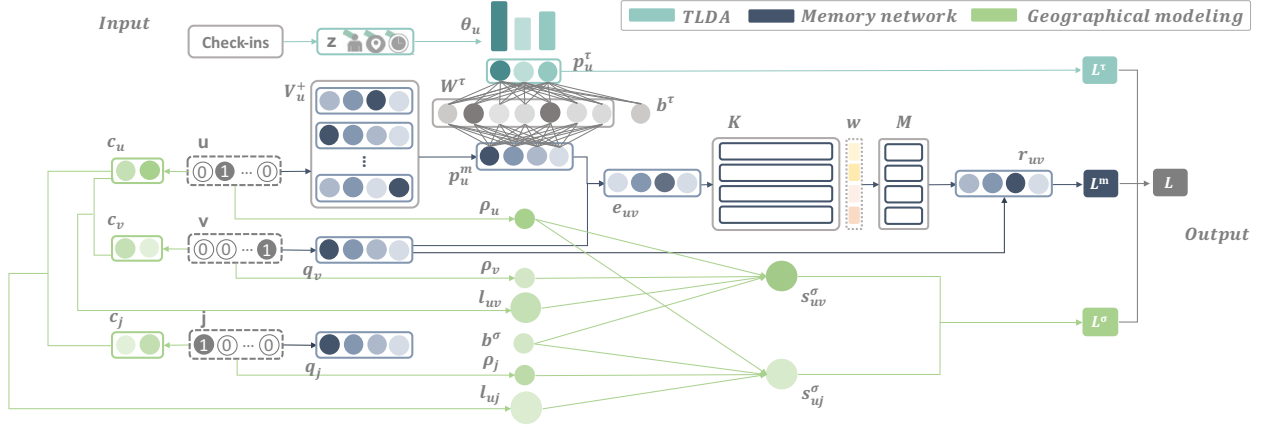
Figure 1: The overall architecture of the TEMN. It consists of three modules represented by different colours. A user, a visited POI, a negative sampling POI, and the user's visiting history are taken as input for POI recommendation.

4.3 Temporal Latent Dirichlet Allocation

Apart from user's memory embedding p_u^m achieved through the memory network, which captures neighbourhood-based interests of a user, we think the intrinsic preferences of a user is also essential in creating a comprehensive user profile for personalized POI recommendation. For this reason, we employ temporal latent Dirichlet allocation (TLDA) [36] in this section to mine the inner interests of users from a more general perspective.

The TLDA model is an unsupervised machine learning algorithm that characterises each user as a mixture of patterns with both venue and temporal preferences considered. In the TLDA model, users' check-in data containing timestamp and POI information are taken as the input. Here the set of a user's historical check-ins as a whole is viewed as a document, where POIs and time slots associated with the check-ins are regarded as words analogously. As a generative probabilistic model in nature, TLDA models the underlying generative process whereby the user's check-in set is created to infer her probability distribution over several latent patterns. In Table 1, the notation and description of the parameters employed in the model to fit observed check-in data are provided. Next we will give a brief explanation about how the intrinsic embeddings of users are estimated during this synthetic process.

To begin with, the pattern distributions of user θ_u and time ϕ_t are sampled from a prior Dirichlet distribution parameterised by α ($\theta_u \sim Dir(\alpha)$) and γ ($\phi_t \sim Dir(\gamma)$), respectively. Then the pattern assignment z_{ut} of each POI can be drawn from a multinomial distribution $z_{ut} \sim Multinomial(\theta_u, \phi_t)$ as below:

$$P(z_{ut}|\alpha, \gamma) = \int_{\theta_u} P(z_u|\theta_u)P(\theta_u|\alpha) \int_{\phi_t} P(z_t|\phi_t)P(\phi_t|\gamma) \quad (8)$$

where z_u is the chosen pattern of user u; and z_t is the chosen pattern of time t. Then a venue v_{ut} is chosen from a multinomial probability φ_z conditioned on the pattern z, which is sampled from a prior Dirichlet distribution β ($\varphi_z \sim Dir(\beta)$) by:

$$P(v_{ut}|z_{ut}) = \int_{\varphi_z} P(v_{ut}|\varphi_z)P(\varphi_z|\beta) \quad (9)$$

Following that, the venue would be generated by estimating the maximum likelihood of v_{ut} by the following equation:

$$P(v_{ut}|\alpha, \gamma, \beta) = \int_{\theta_u} \int_{\phi_t} \int_{\varphi_z} P(v_{ut}, z_{ut}, \theta_u, \phi_t, \varphi_z|\alpha, \gamma, \beta) \quad (10)$$

Eventually, the probability distribution over π patterns for a given user u can be estimated by applying the Gibbs sampling algorithm. It can be seen as a pattern mixture that represents each user's intrinsic embedding denoted as p_u^τ:

$$p_u^\tau = (P(z_u^1|\theta_u), P(z_u^2|\theta_u), \cdots, P(z_u^\pi|\theta_u)) \quad (11)$$

It is worth noting that the TLDA module not only provides us the pattern-user probability distribution, the distributions of venues and time slots associated with each pattern can also be estimated to help us understand what each pattern might refer to.

Notation	Description
α	Dirichlet prior over the pattern-user distributions
β	Dirichlet prior over the venue-pattern distributions
γ	Dirichlet prior over the pattern-time distributions
θ_u	pattern distribution of user u
φ_z	venue distribution of pattern z
ϕ_t	pattern distribution of time t
z_{ut}	pattern of venue of user u at time t
v_{ut}	venue of user u's check-in at time t

Table 1: Notation and description of parameters in TLDA.

4.4 Fusion of MN and TLDA

In the last two subsections, the modules of memory network and TLDA have been introduced, respectively. Essentially, they are not two separate parts, but an interrelated whole system that is capable of learning both neighbourhood-based and global characteristics of users and POIs, as well as the fine-grained relationships between specific user-POI pairs. In this subsection, we will introduce how the memory network and TLDA can be fused together, and mutually reinforce each other under the TEMN framework.

The key part connecting the two modules is user embedding. More specifically, we define our problem as a multi-class classification task and use a neural network approach to learn the topic-enhanced user embedding for the memory network. As our main aim is to map the memory embedding p_u^m of a user u to her intrinsic embedding p_u^τ through the neural network architecture, p_u^m is taken as the input data, and the intrinsic embedding p_u^τ of the user is applied as the output where each element of the vector represents a class. Mathematically, the model is defined as:

$$O^\tau = f(W^\tau p_u^m + b^\tau) \qquad (12)$$

where W^τ and b^τ denote the weight matrix and bias vector, respectively. $f(\cdot)$ is the activation function. To get the final predicted probabilities $\hat{p}_u^\tau$, the output vector O^τ is converted by the softmax function as:

$$\hat{p}_u^\tau = softmax(O^\tau) \qquad (13)$$

By introducing softmax activation function in the output layer, the cross-entropy loss is used for optimisation:

$$\mathcal{L}^\tau = -\sum_u \sum_i p_u^\tau(i) \cdot log(\hat{p}_u^\tau(i)), \forall i = 1, 2, \cdots, \pi \qquad (14)$$

where i represents one of the π patterns in the TLDA. The selection of activation function $f(\cdot)$ in Equation 12 is not necessarily restricted to a particular function. Many options like sigmoid, hyperbolic tangent, and ReLU can be taken in practice according to specific application domain. Alternatively, one can also choose to apply Equation 12 without an activation function and convert it to a softmax regression model. Here, we choose the ReLU function as it yields the most favourable result.

4.5 Geographical Modeling

Given the TEMN is expected to be competent for context-aware recommendation, it is necessary to incorporate spatio-temporal information into the model. As introduced in the last subsection, temporal effects have been considered in the design of TLDA. In this subsection, we are going to present how geographical influence is taken into account in the TEMN.

Most existing methods assume that the geographical effect between a pair of user and POI is determined by their physical distance, which fails to capture the asymmetry between the user and POI, the high variation of geographical preferences across users, and the differences of geographical influence among POIs. In this sense, we think that to measure the geographical effect only with physical distance is not enough. Instead, we introduce the concepts of *geographical preference* and *geographical influence* which

are specifically learned for each user and POI so as to depict the fine-grained geographical effect between them.

For a user u, each POI $v \in \mathcal{V}_u^+$ visited by her has certain geographic coordinates $c_v = (g_{lat}^v, g_{lon}^v)$ recorded simultaneously when the certain check-in was created. Here g_{lat}^v and g_{lon}^v represent the latitude and longitude of venue v, respectively. Through an investigation about her check-in history, the central point c_u of user u where her check-ins clustered around geographically can be determined by:

$$c_u = (\frac{1}{|V_u^+|} \sum_i g_{lat}^i, \frac{1}{|V_u^+|} \sum_i g_{lon}^i), \forall i \in |V_u^+| \qquad (15)$$

Then the physical distance l_{uv} between user u and venue v can be calculated through:

$$l_{uv} = ||c_u - c_v||_2 \qquad (16)$$

To model the geographical effect between each user-POI pair $<u, v>$, three factors are introduced in this part: 1) the *geographical preference* ρ_u of user u, which characterises how sensitive the user is to distance; 2) the *geographical influence* ρ_v, representing the influential level of the venue reflected on physical distance; and 3) the geographical distance between them l_{uv}. Based on this, the geographical score s_{uv}^σ of user-POI pair $<u, v>$ is calculated by:

$$s_{uv}^\sigma = \rho_u l_{uv} + \rho_v l_{uv} + b_u^\sigma = (\rho_u + \rho_v)l_{uv} + b^\sigma \qquad (17)$$

where b^σ denotes the geographical bias. Similarly as that in memory network (Section 4.2), the objective function applied here is also based on BPR[21]. Given a venue v which has been visited by user u, and a negative venue j that she never interacted with, the pairwise loss function is defined as:

$$\mathcal{L}^\sigma = \sum_u \sum_{v \in V_u^+} \sum_{j \notin V_u^+} max(0, s_{uj}^\sigma - s_{uv}^\sigma + \lambda^\sigma) \qquad (18)$$

where λ^σ is the margin separating positive and negative examples.

4.6 Joint Training

Our complete recommender system is a hybrid of memory network, TLDA, and geographical modeling, which is a topic-enhanced memory network with spatio-temporal contextual information considered. Until now, we have elaborated each module respectively and introduced how memory network is linked with the TLDA. In this part, we present how the TEMN framework functions as a whole.

For each pair of user-POI $<u, v>$, we can calculate its score achieved from the memory network s_{uv}^m through Equation 6, and its geographical score s_{uv}^σ by Equation 17. Then the overall score s_{uv} can be easily obtained through the following equation:

$$s_{uv} = s_{uv}^m + \eta s_{uv}^\sigma \qquad (19)$$

where η is a weighting parameter, which can be optimally tuned to study the effects of incorporating geographical modeling for POI recommendation under our framework. The overall objective function for TEMN is a weighted sum of all the contributions of each related individual objective as:

$$\mathcal{L} = \mathcal{L}^m + \varsigma \mathcal{L}^\tau + \epsilon \mathcal{L}^\sigma + \lambda \|\varrho\|_2^2 \qquad (20)$$

where $\mathcal{L}^m$, $\mathcal{L}^\tau$, and $\mathcal{L}^\sigma$ denote the loss induced by the memory network, TLDA, and geographical modeling respectively. ς and ϵ are weighting parameters specified to influence to what degree the TLDA and geographical effects should be taken into account during the optimisation. ϱ is the model parameter set; and λ is a parameter which controls the importance of the last term, where we regularise all the model parameters to prevent overfitting. In the training phase, mini-batch stochastic gradient descent (SGD) is employed to minimise the objective function.

5 EXPERIMENTS

The effectiveness of the proposed model for top-N POI recommendation under various scenarios is evaluated in this section. We begin with experimental setup, and then present and analyse the experimental results both quantitatively and qualitatively.

5.1 Experimental Setup

5.1.1 Datasets. We conduct our experiments on the WeChat[3] Moments check-in data. This study was approved by Tencent[4] and the Computer Science and Technology Department University of Cambridge Ethics Committee. Initially, the whole dataset contains all the POI check-in records created by WeChat users living in Beijing for one year spanning from September 2016 to August 2017. After filtering the data, we obtain an evaluation dataset, termed as *WeChat(GPR)* for general and context-aware recommendation. For sequential POI recommendation, it is further processed as that introduced in Section 3. This newly created dataset containing the set of successive check-ins for each user is named as *WeChat(SPR)*. The statistics of the two datasets are summarised in Table 2. Apart from the interaction information between users and POIs, the timestamp associated with each check-in and the side information of each POI consisting of coordinates and categories (e.g., restaurant, art museum, sports centre) are also available.

Dataset	#Users	#POIs	#Chek-ins	Density
WeChat (GPR)	75,973	28,183	564,4965	0.264%
WeChat (SPR)	28,566	13,826	509,589	0.129%

Table 2: Statistics of the evaluation datasets.

5.1.2 Evaluation Metrics. Two common ranking evaluation metrics, *Hit Ratio (HR)* and *Normalised Discounted Cumulative Gain (NDCG)* are adopted. HR measures whether the test POI shows within the top N in the ranked list, and the NDCG takes the position of the test POI into account and penalises the score if it is ranked lower in the list.

5.1.3 Baselines. We compare our proposed approach against competitive baselines representing latent factor models, topic models, metric learning methods, Markov chain models, and deep learning-based models under various recommendation scenarios.

To evaluate the performance of our model in general POI recommendation, the following methods are considered:

- **MF** [15] Matrix factorization is one of the most popular model-based collaborative filtering approaches for recommendation, which models the user-POI relationship using inner product.
- **BPR** [21] Bayesian personalised ranking optimises the MF model with a pairwise ranking loss. It is tailored to recommendations with implicit feedback data.
- **LDA**[1] It is an unsupervised learning approach initially applied to cluster documents and discover topics based on their contents. It is the prototype of our TLDA module that does not consider temporal factors in human mobility.
- **CML** [13] This is a method used to exam whether the latent relational learning of r_{uv} is necessary in our design of memory network module. CML minimises the Euclidean distance between user and POI vectors as $||p_u^m - q_v||_2^2$.
- **LRML** [23] This is a memory network-based approach which learns a translation vector between a pair of user and item. It can be seen as a memory network module without topic-enhanced effects and geographical influence.
- **TEMN(GPR)** It is a variant of TEMN, which reserves the memory network component, replaces the TLDA with LDA, and removes geographical modeling part.

The second group of baselines are models devised for sequential POI recommendation, which include:

- **LORE** [32] It considers sequential influence and geographical effects in recommendation by adopting additive Markov chain.
- **ST-RNN** [18] This is a RNN-based state-of-the-art POI recommendation model, which considers information about the time interval and distance between pair of check-ins.
- **TEMN(SPR)** Our complete model with the WeChat (SPR) data fed into, which is employed to test our model's effectiveness in successive POI recommendation.

The last group of baseline approaches supporting the integration of contextual information associated with POI check-ins are:

- **GeoMF** [17] This method extends the MF by augmenting latent factors with the user's activity region and POI's influence area.
- **TLDA** [36] TLDA is an extension model of LDA that can be applied to extract users' lifestyle patterns with temporal preferences considered. It is also employed as the topic module in our framework to enhance memory network.
- **TEMN(CPR)** It is our complete model with the WeChat (GPR) dataset employed.

5.2 Experimental Results

The comparison results of our proposed model and the baselines on two WeChat datasets in three POI recommendation scenarios are reported in Table 3. The key observations from the experimental results are summarised as follows:

(1) Encouragingly, it is clear that the performance of our proposed model TEMN is consistently better than all the baselines under different conditions by a relative large margin (the relative improvement ratio over the best baseline in the general, context-aware, and sequential recommendation scenario is 5.16%, 3.25%, and 27.95%, respectively). Additionally, as can be seen from both Table 3 and Figure 2, TEMN offers significantly larger improvement for NDCG than HR, suggesting that the proposed model gives a higher ranking to the positive test POI, and shows the feasibility

Type	Method	HR@5	NDCG@5	HR@10	NDCG@10
General	MF	0.50961	0.30345	0.63174	0.34028
	BPR	0.52561	0.33747	0.64313	0.37597
	LDA	0.69221	0.53960	0.80958	0.57778
	CML	0.67983	0.45613	0.78813	0.49163
	LRML	0.69594*	0.54669*	0.81199*	0.58436*
	TEMN(GPR)	**0.70389**(+1.141%)	**0.55221**(+1.010%)	**0.81752**(+0.682%)	**0.58914**(+0.817%)
Context-aware	GeoMF	0.63714	0.50435	0.70742	0.52741
	TLDA	0.71518*	0.55852*	0.83033*	0.59601*
	ST-RNN	0.60240	0.47044	0.74372	0.51621
	TEMN(CPR)	**0.72876**(+1.899%)	**0.57666**(+3.248%)	**0.83398**(+0.440%)	**0.61053**(+2.436%)
Sequential	LORE	0.53187*	0.37880	0.69148*	0.43049
	ST-RNN	0.53041	0.42867*	0.62798	0.45999*
	TEMN(SPR)	**0.62105**(+16.767%)	**0.54847**(+27.947%)	**0.69769**(+0.897%)	**0.57304**(+24.577%)

Table 3: Performance comparison of different methods in three recommendation scenarios. Best performance is in boldface. We use "*" to mark the best performance from baselines for each comparison and report the improvement ratio of our model over the best baseline performance for each scenario in parentheses.

and effectiveness of applying the topic-enhanced memory network to the top-N POI recommendation.

(2) Considering the three variants of the TEMN, we find that the overall performance order is as follows: TEMN(GPR) > TLDA> LRML. This indicates that an integrated structure (without contextual information) of memory network and topic model performs better than the individual components. To further compare the performance of memory network module (LRML) and topic model module (TLDA), TLDA shows its superiority by giving around 2-3% improvement across all the metrics and cut offs. TLDA is a model proposed by us last year to discover individuals' lifestyle patterns. This is the first time it is applied to the recommendation task and provides stunning performance by beating all the other baselines. In a nutshell, even though the performance of the individual modules is marginally worse than the complete TEMN model, they provide the best baseline performance in general and context-aware recommendation scenario, respectively. Moreover, the outstanding performance of TLDA highlights the importance of users' long-term interests mining and temporal effect consideration in POI recommendation.

(3) Taking other baseline methods for the first two recommendation scenarios into account, neural network-based approaches (CML, LRML, and ST-RNN) and topic models (LDA and TLDA) outperform MF-based methods (MF, BPR, and GeoMF) in most cases, which indicates the usefulness of capturing more complex user-POI relationships through nonlinear methods and the global preference of users through topic modeling. The relatively poor performance of MF-based approaches suggests that adopting dot product may not be enough to depict user-POI interactions as discussed in [12]. Another interesting finding is that LRML performs better than CML, proving the advantage of employing a translation vector to capture relationships between a user-POI pair. In addition, GeoMF, TLDA, and TEMN (CPR) all give better recommendation results than their prototypes (MF, LDA, and TEMN(GPR)), indicating the value of incorporating spatio-temporal information into POI recommender

systems. The unfavourable performance of ST-RNN may be due to RNN-based models' problem of dealing with long sequences [14]. It is a more suitable approach for successive POI prediction, where recent visits are particularly important.

(4) Utilising a different dataset (WeChat (SPR)), the experimental results of models in the third group for sequential POI recommendation is not directly comparable with that in the other two groups and is discussed individually here. As it can be seen from the table, our proposed model TEMN(SPR) gives the best performance again with dramatic improvement ratios in most cases. These results indicate the effectiveness of TEMN for sequential recommendation, which is actually not surprising because the memory mechanism as well as the geographical modeling design in our architecture provide better expressive power to model recent user-POI interactions and fine-grained geographical factors for each specific user-POI pair. Among the baseline methods applied to sequential POI recommendation task, it is interesting to see that Markov chain model (LORE) performs better for HR, while RNN-based method (ST-RNN) shows better results for NDCG. This finding can be utilised in practical application that when the length of recommendation list is limited (N is small in the top-N recommendation), our TEMN model and ST-RNN are better choices for successive POI recommendation.

5.3 Detailed Analysis of the Proposed Model

After discussing the experimental results quantitatively, in this subsection we present some qualitative observations and analysis results of TEMN.

5.3.1 Attention Visualisation. As introduced in Section 4, the proposed architecture of TEMN incorporates topic information and attention mechanism in the memory network, which enables us to visualise the weighted importance of memory slices with respect to multiple patterns and thus enhance the interpretability of our model. We plot a heatmap of the weights in Figure 3, where the colour scale represents the strength of the attention weights, and
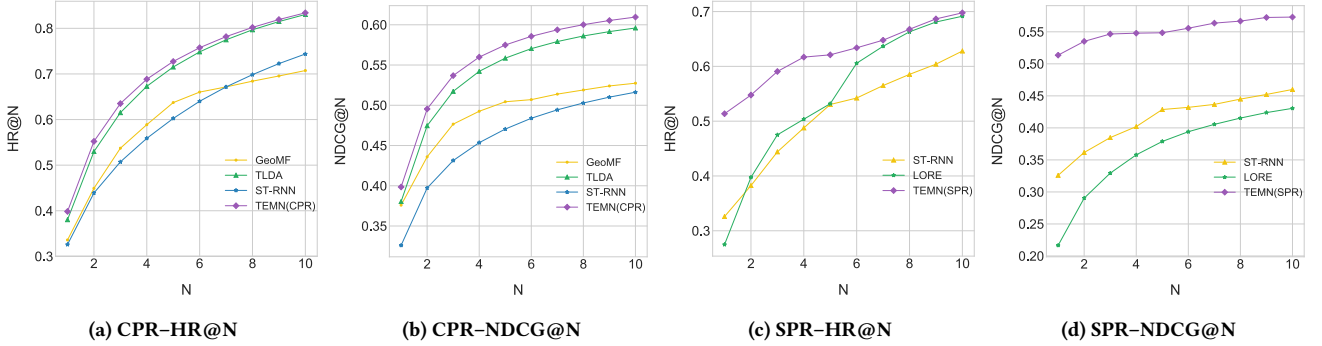
(a) CPR–HR@N   **(b) CPR–NDCG@N**   **(c) SPR–HR@N**   **(d) SPR–NDCG@N**

Figure 2: Evaluation of Top-N POI recommendation where N ranges from 1 to 10 on the two WeChat datasets in two scenarios.

each row represents the average attention vector for each pattern. As can be observed, it is clear that each pattern has learned a certain type of selection rules across memory slices. Moreover, some patterns are particularly similar with each other than their counterparts. When we compared these results with the semantic meaning for each pattern as presented in Table 4, we can find that similar patterns like 2, 7, 8, and 9 all have residential district as their top category. In contrast, pattern 6, which is related to Olympic venues, shows a distinctive attention vector, suggesting that the visiting features of this pattern can be extremely different from other types.

Pattern	Top Categories
0	Airport, Central business district, Hub station
1	Shopping mall, Square, Commercial street
2	University, Shopping mall, Residential district
3	Bar street, Stadium
4	Historic district, Shopping mall
5	High-tech zone, University, Shopping mall
6	Olympic Park
7	Residential district, SOHO
8	Hub station, Residential district
9	University, Residential district

Table 4: Pattern and top categories.

5.3.2 Geographical Preferences of Patterns. As introduced in 4.5, the *geographical influence* ρ_v of a POI and *geographical preference* ρ_u of a user can be learned through our model. By calculating the average *geographical influence* of POIs and *geographical preference* of users for each pattern, we can gain insight into the geographical characteristics of different patterns. The calculated results are presented in Figure 4, where a smaller number means the POI or user is less sensitive to distance. It can be seen that patterns of 0 and 8 give the smallest value of ρ_v, indicating that the visits to POIs belonging to these patterns (e.g. airport, central business district, hub station) are not influenced by distance too much; while for pattern 1, which is represented by POI categories more related to local
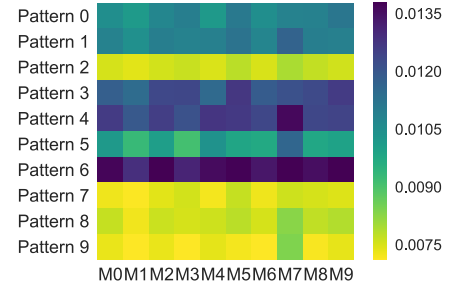

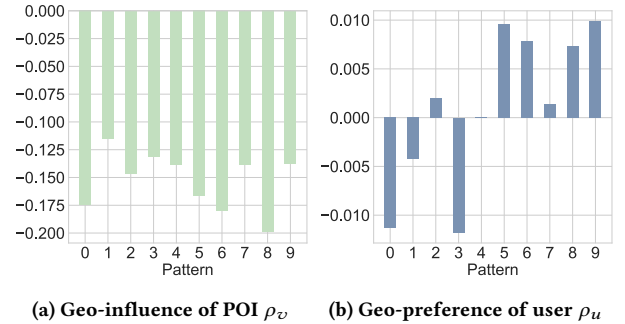**Figure 3: Attention weights over memory slices for patterns.**

(a) Geo-influence of POI ρ_v   **(b) Geo-preference of user ρ_u**

Figure 4: Average geographical influence of POIs and geographical preference of users for patterns.

services, it is rather sensitive to distance that people are unlikely to visit them from distant places. On the other hand, users from pattern 0 and 3 are relatively more willing to visit POIs far away than their counterparts by showing the smaller average ρ_u.

6 CONCLUSION

We introduce a novel end-to-end architecture, named Topic-Enhanced Memory Network (TEMN), for better POI recommendation with spatio-temporal contextual information considered. Different from

existing deep learning frameworks for POI recommendation, TEMN is a unified hybrid model that leverages TLDA and external memory network with a neural attention mechanism to capture both global and fine-grained preferences of users. Comprehensive experiments under multiple configurations demonstrate the proposed architecture's significant improvements over competitive baselines. Qualitative visualisation of the attention weights and semantic investigation about patterns learned via TLDA module provide insight into the learning process and the recommendation results. In future work, we plan to extend TEMN to incorporate richer content information such as users' profiles and POIs' attributes.

ACKNOWLEDGMENTS

REFERENCES

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.
[3] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-Aware Hierarchical Point-of-Interest Embedding Model for Successive POI Recommendation.. In *IJCAI*. 3301–3307.
[4] Meng Chen, Yang Liu, and Xiaohui Yu. 2014. Nlpmm: A next location predictor with markov modeling. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 186–197.
[5] Chen Cheng, Haiqin Yang, Irwin King, and Michael R Lyu. 2012. Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks.. In *Aaai*, Vol. 12. 17–23.
[6] Chen Cheng, Haiqin Yang, Michael R Lyu, and Irwin King. 2013. Where You Like to Go Next: Successive Point-of-Interest Recommendation.. In *IJCAI*, Vol. 13. 2605–2611.
[7] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1082–1090.
[8] Travis Ebesu, Bin Shen, and Yi Fang. 2018. Collaborative Memory Network for Recommendation Systems. *arXiv preprint arXiv:1804.10862* (2018).
[9] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. 2017. POI2Vec: Geographical Latent Representation for Predicting Future Visitors.. In *AAAI*. 102–108.
[10] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2013. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 93–100.
[11] Jing He, Xin Li, and Lejian Liao. 2017. Category-aware next point-of-interest recommendation via listwise Bayesian personalized ranking. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 1837–1843.
[12] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.
[13] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 193–201.
[14] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. 2018. Improving sequential recommendation with knowledge-enhanced memory networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 505–514.
[15] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.
[16] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. 2016. Point-of-interest recommendations: Learning potential check-ins from friends. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 975–984.
[17] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. 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*. ACM, 831–840.
[18] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts.. In *AAAI*. 194–200.
[19] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 739–748.
[20] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. 2018. A Contextual Attention Recurrent Architecture for Context-Aware Venue Recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 555–564.
[21] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.
[22] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. 2013. Exploiting local and global social context for recommendation.. In *IJCAI*, Vol. 13. 2712–2718.
[23] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Latent Relational Metric Learning via Memory-based A ention for Collaborative Ranking. In *Proceedings of the 27th International Conference on World Wide Web*.
[24] Hao Wang, Yanmei Fu, Qinyong Wang, Hongzhi Yin, Changying Du, and Hui Xiong. 2017. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1135–1143.
[25] Hao Wang, Huawei Shen, Wentao Ouyang, and Xueqi Cheng. 2018. Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation.. In *IJCAI*. 3877–3883.
[26] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 391–400.
[27] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 15–24.
[28] Mao Ye, Xingjie Liu, and Wang-Chien Lee. 2011. Exploring social influence for recommendation-a probabilistic generative model approach. *arXiv preprint arXiv:1109.0758* (2011).
[29] Hongzhi Yin, Weiqing Wang, Hao Wang, Ling Chen, and Xiaofang Zhou. 2017. Spatial-aware hierarchical collaborative deep learning for POI recommendation. *IEEE Transactions on Knowledge and Data Engineering* 29, 11 (2017), 2537–2551.
[30] Chao Zhang, Keyang Zhang, Quan Yuan, Luming Zhang, Tim Hanratty, and Jiawei Han. 2016. Gmove: Group-level mobility modeling using geo-tagged social media. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1305–1314.
[31] Jia-Dong Zhang and Chi-Yin Chow. 2015. GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 443–452.
[32] Jia-Dong Zhang, Chi-Yin Chow, and Yanhua Li. 2014. Lore: Exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 103–112.
[33] Kaiqi Zhao, Gao Cong, Quan Yuan, and Kenny Q Zhu. 2015. SAR: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 675–686.
[34] Shenglin Zhao, Tong Zhao, Irwin King, and Michael R Lyu. 2017. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th international conference on world wide web companion*. International World Wide Web Conferences Steering Committee, 153–162.
[35] Meizi Zhou, Zhuoye Ding, Jiliang Tang, and Dawei Yin. 2018. Micro behaviors: A new perspective in e-commerce recommender systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 727–735.
[36] Xiao Zhou, Anastasios Noulas, Cecilia Mascoloo, and Zhongxiang Zhao. 2018. Discovering Latent Patterns of Urban Cultural Interactions in WeChat for Modern City Planning. *arXiv preprint arXiv:1806.05694* (2018).
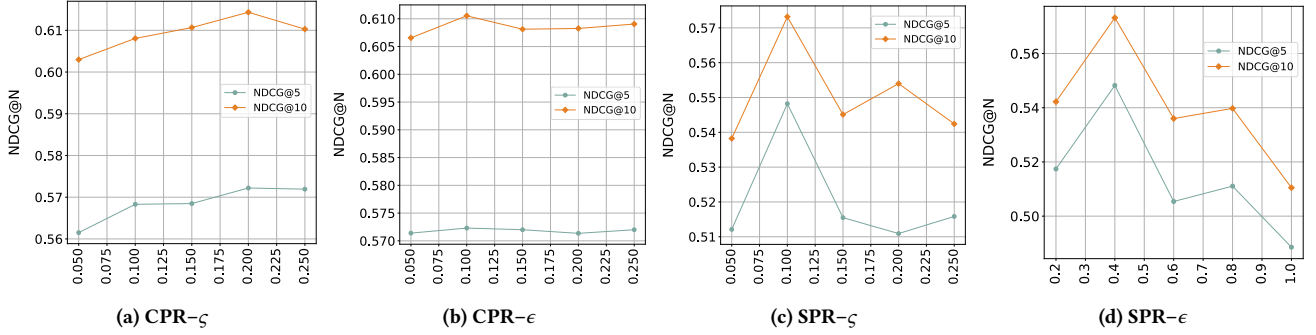
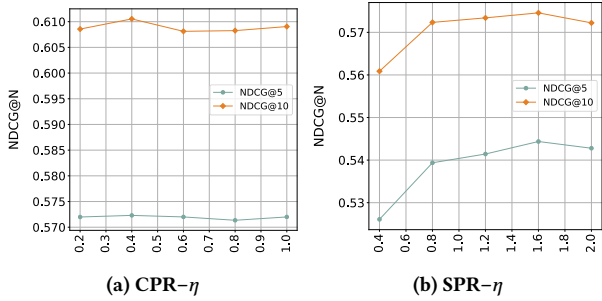**Figure 5: Effect of weighting parameters in the loss function.**

Figure 6: Effect of weighting parameter in score function.

A APPENDIX

In this section, we provide detailed explanations of how we did our experiments for reproducibility. Our code is publicly available at https://github.com/XiaoZHOUCAM/TEMN.

A.1 Dataset Filtering

To ensure the quality of the data, we retain users who interacted with at least ten unique POIs. The filtered dataset is termed as *WeChat(GPR)* and employed as the evaluation dataset for both general and context-aware recommendation. For sequential POI recommendation scenario, the dataset is further processed as that introduced in *Sequential POI recommendation*, Section 3. More specifically, we set the threshold ΔT to one day (24 hours), and keep sequences with at least five successive check-ins. This newly created dataset containing the set of successive check-ins for each user is called *WeChat(SPR)*.

A.2 Evaluation Protocols

In the experiments, the check-in data are transformed into binary implicit feedback as ground truth, indicating whether the user had visited the specific POI. For each dataset, we holdout the latest 15% visiting history of each user to construct the test set, and split the remaining data into training (70%) and validation (15%) sets. The validation set is used to tune hyper-parameters and the final performance comparison is conducted on the test set. Closely

following the setup that in [12, 21], we randomly sample 100 POIs that are not visited by the user and rank the test POI along with the 100 negative samples for evaluation.

A.3 Parameter Tuning

We implemented the TEMN model in TensorFlow[5]. For our model, all hyperparameters are tuned according to the validation set based on the NDCG metric. We randomly initialise model parameters according to the uniform distribution and optimise the model by conducting stochastic gradient descent (SGD). The learning rate of SGD and the regularisation parameter are determined by grid search in the range of {0.0001, 0.005, 0.001, 0.01} and {0.0001, 0.001, 0.01, 0.1}, respectively. The dimensionality of user and item embeddings d is tuned amongst {20, 50, 75, 100}. The number of memory slices in M is tuned amongst h = {5, 10, 20, 50}. And the number of patterns set in TLDA is tuned among {3, 5, 10, 20}. For MN module and geographical modeling that minimise the hinge loss, the margin is tuned amongst {0.1, 0.2, 0.25, 0.5}.

A.4 Effect of Weighting Parameters

The weighting parameters utilised in the overall loss function (Equation 20) and score function (Equation 19) are tuned by us to study the effect of incorporating TLDA module and geographical modeling for POI recommendation. Specifically, we study the performance of our models on NDCG@5 and NDCG@10 in context-aware POI recommendation (CPR) and sequential POI recommendation (SPR) scenarios by tuning the relevant weighting parameters as shown in Figure 5 and Figure 6.

It can be seen from Figure 5a and Figure 5c, when the weighting parameter of TLDA component ς is set to a small value of 0.05, models for CPR and SPR all give unfavourable results. When TLDA begins to play an increasingly important role, the performance is improved dramatically. The best results are achieved when $\varsigma \approx 0.2$ for CPR and $\varsigma \approx 0.1$ for SPR. However, when the weights of TLDA module continue to rise, the performance goes down. This result indicates that integrating TLDA module with memory network indeed helps to offer better POI recommendation, however, paying

---

[5]https://www.tensorflow.org/

too much attention to user's intrinsic preference may weaken her neighbourhood-based characteristics.

As for the influence of geographical modeling, an interesting observation is that in CPR scenario, the performance of models is stable when ϵ is tuned; while for SPR, the curves fluctuate substantially with the increasing of ϵ. Moreover, it can be found that when the weighting parameters of geographical module are set to relatively larger values ($\epsilon \approx 0.4$ and $\eta \approx 1.6$), the performance of our models is most satisfying. These findings suggest that geographical effects have a greater influence on sequential POI recommendation, where shor-term preferences value more.

A.5 Parameter settings

Eventually, we find that the following hyperparameters work well: learning rate is set to 0.005, regularisation parameter is set to 0.0001, both of λ^m and λ^σ are set to 0.2. The dimension of user and item embeddings d is set to 50, and the number of memory slots h is set to 10. The number of patterns set in TLDA is 10. We set the weighting parameters $\varsigma = 0.2$, $\epsilon = 0.1$, and $\eta=0.4$ in general and context-aware recommendation; while for sequential POI recommendation, we set $\eta = 1.6$ in the overall score function, ς and ϵ to 0.1 and 0.4, respectively.