Applied Intelligence (2021) 51:1904–1917
https://doi.org/10.1007/s10489-020-01868-0

Exploiting multi-attention network with contextual influence for point-of-interest recommendation

Liang Chang[1,2] · Wei Chen[3] · Jianbo Huang[1] · Chenzhong Bin[1] · Wenkai Wang[1]

Accepted: 6 August 2020 / Published online: 17 October 2020

Abstract

Point-of-Interest (POI) recommendation has become an important service on Location-Based Social Networks (LBSNs). In order to improve the performance of recommendation, besides the check-in data generated in LBSNs, researchers are striving to exploit various auxiliary information such as social relation among users and geographical influence among neighbourhood POIs. However, existing works cannot effectively study the diverse degrees of influence from user's friends, neither are they able to capture the feature impacts of POIs in the preference modelling process. To overcome these challenges, by making use of a **M**ulti-**A**ttention **N**etwork to learn the **C**ontextual influence of both users and POIs, this paper presents a model named MANC for POI recommendation. The MANC model consists of two parts: a user-friend module and a POI neighbourhood module. Unlike existing works which treat the influences from different friends of a user equally, the user-friend module in MANC applies an attention-based memory component to generate specific relation vectors which can differentiate the influence from the aspect of interest, and applies a friend-level attention network to adaptively capture the preferences of users. For the POI contextual information, the POI neighbourhood module in MANC applies a feature-level attention network to capture the latent features of neighbourhood POIs, and applies a POI-level attention network to capture the geographical influence among POIs. Extensive experiments are carried out, and it is shown that the MANC model achieves better performance than other state-of-the-art methods.

Keywords Point-of-interest · Recommendation system · Attention network · Collaborative filtering · Contextual information

1 Introduction

Advancements of mobile internet technologies have led to the emergence of location-based social networks (LBSNs) services such as Foursquare, Yelp, and Facebook [22, 25, 27]. More and more individuals are sharing their experience of visiting places with their friends via LBSNs. The large amount of check-in data generated in LBSNs can be used for the recommendation of places such as cafes and museums which are called Point-of-Interests (POIs). POI recommendation plays an important role in helping visitors to find interesting places. However, different from traditional e-commerce recommendation, POI recommendation often suffers the problem of sparse data. Furthermore, the check-in data of users for POI recommendation is implicit data rather than explicit ratings (e.g., 1-5). As a result, traditional collaborative filtering (CF) methods such as matrix factorization (MF) [16] and its various extensions [11, 19, 20] are unable to perform well for POI recommendation.

To alleviate the data sparse problem, many POI recommendation methods try to make use of auxiliary information related with check-ins, such as the temporal signal [45], sequential dependence [41, 44], social relation among users [18, 35, 38], and geographical distance between POIs [19, 20, 24]. Among these auxiliary information, it is shown that social relation and geographical distance are two of the most important factors that affect user's decisions [25]. However, existing methods that make use of these two kinds of information still have the following limitations.

One limitation is that, when making use of social relation, existing methods [18, 35] assign the same weight to all friends of a user, and therefore cannot consider different degrees of influence coming from different friends of the

✉ Wei Chen
w_chen369@163.com

Extended author information available on the last page of the article.

same user. As we know, given a user and one of his or her friends, there are often some common preferences existing between them. A user's preferences can often be inferred from the preferences of his or her friends. Furthermore, given a user, he or she might have different common preferences with different friends. As an example, in Fig. 1. both User A and his or her friend User B like drinking coffee, while the common preference of User A and his or her friend User C is reading books. Therefore, if User A wants a recommendation on cafes, the weight of advice from User B is likely to be higher than that from User C. Correspondingly, if User A wants to visit a bookshop, the advice from User C is more valuable. In addition, in reality, a user's friends often have different and dynamic influences on the user. For example, a user who never goes to cinema before will adopt the suggestion of his or her friends who often go to cinema, but considering traveling, he or she will turn to those friends who enjoy traveling.

Another limitation is that, when making use of geographical distance, existing methods [24, 27, 32] cannot capture the feature influence across POIs. Generally, individuals have different preferences for the characteristics of POIs [40]. For example, in the same region, some tourists prefer to visit historic buildings while others enjoy seeing natural attractions. According to the feature correlations of visited and unvisited POIs for each user, we can infer the relevant score between a user and the POIs that have not been visited.

Motivated by the above observations, we present a novel model named MANC that utilizes a Multi-Attention Network to learn the Contextual influence of both users and POIs for POI recommendation. The MANC model consists of two integral components, namely the user-friend module and the POI neighbourhood module. The user-friend module first captures the influence on user's interest aspect by applying an aspect-level attention network, and subsequently models the influence on the user's overall preferences by employing a friend-level attention network. In the POI neighbourhood module, a feature-level attention network is initially applied to learn the specific vectors for describing the correlation of a POI's latent features among neighbourhood POIs. A POI-level attention network is then applied to model the geographical influence through a distance decay function, such that unvisited POIs that are more similar and closer to the check-in POIs can be visited by users with greater probability. Comparisons with existing methods on benchmark datasets demonstrate the improved performance of our model.

In summary, the major contributions of this paper are as follows.

1) To differentiate the influence of a user's friends on the preferences of the user, we develop an aspect-level attention network component to calculate the differences in interest in the user-friend common preferences and exploit an additional attention network to model the effect of the user's overall preferences resulting from the user's friends.
2) To incorporate the POI's neighbouring influence, we design a POI neighbourhood module that introduces two attention networks to model the influence of geographical distance and the correlation of the POI's latent features.

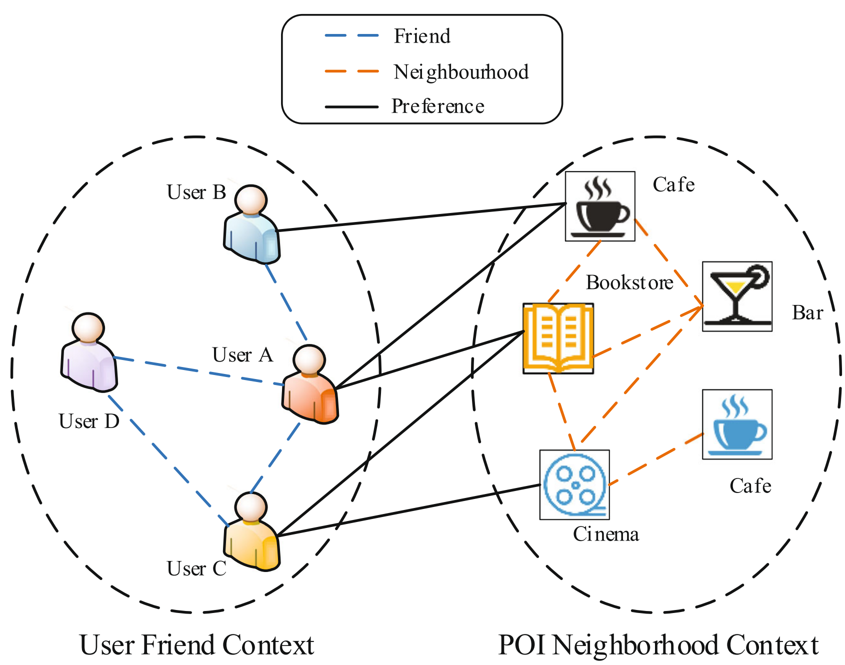


Fig. 1 Example of user preferences in social relationships. User A and User B have a preference for the Cafe while User A and User C have a preference for the Bookstore. The common preference aspects of these two pairs are distinct

3) We carry out comprehensive experiments to evaluate the performance of the recommendation. Results demonstrate that the performance of the proposed model surpasses other state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 presents relevant definitions and the problem statements. Our method is presented in Section 4, which is followed by an experimental evaluation in Section 5. Section 6 concludes this paper.

2 Related work

2.1 POI recommendation

The rapid progress of LBSNs in recent years has resulted in POI recommendation being the focal point of much attention. The majority of the relevant research is based on the application of CF with user's historical check-in data for personalized location recommendations [36–38]. The most effective CF is based on matrix factorization and its variants. These methods convert the check-in matrix into a low-dimensional user matrix and a low-dimensional POI matrix, and apply the inner product to predict unvisited POIs that users may be potentially interested in. Furthermore, several studies have considered user check-in records as implicit feedback that can be modelled by applying the weighted regularized matrix factorization [18, 20], while others found that the recommendation problem can be treated as a pairwise ranking task [23, 44]. Liu et al. [23] adopted an adaptive Bayesian Personalized Ranking (BPR) loss [30] to optimize the rankings for POI recommendation.

To improve the recommendation efficiency, researchers integrated rich contextual information into their proposed models [25]. Some researchers incorporated social relations to infer users' preferences [1, 3, 18]. Such methods are based on explicit social relations [2] and are generally considered to improve accuracy. For example, Li et al. [18] modelled three types of social correlations (i.e., social friends, location friends and neighbouring friends) to learn the influence of friend check-ins to the user. Moreover, others have utilized geographical influence to recommend POIs. [21, 29] Hossein et al. [29] modelled geographical influence via the relevance of each location to the user's principle region of activity. In addition, Zhang et al. [43] presented a special three-layer network to model multi-tag, social and geographical influences.

Recent studies have applied deep neural networks for POI recommendations. For example, Yang et al. [35] designed a neural collaborative filtering method to model complex user-POI interactions using user and POI contexts. Chen et al. [23] presented SAE-NAD, a high performance model amongst the current literature. In particular, SAE-NAD applied a self-attention auto-encoder with neighbor-aware influence. Our work is distinct from SAE-NAD in the following two aspects: (1) SAE-NAD only considers the influence of the distance among neighbourhood POIs, while our method which is based on the distance influence, also considers the correlation of latent features across POIs; and (2) we introduce the user-friend context to capture the different influences on user preferences.

2.2 Deep learning in recommendation

The application of deep learning has been successful in numerous fields, including natural language understanding [4], image processing [9] and speech recognition [15]. Multiple researchers have also tried to improve recommendation performance via deep learning. He et al. [14] designed a Neural Collaborative Filtering (NCF) method that learns complex user-item interactions using a deep neural network, with the ability to outperform numerous traditional approaches. Further to this, Zheng et al. [46] proposed a deep cooperative neural model for review recommendation.

The attention mechanism has been demonstrated as an effective tool in many deep learning tasks, such as machine translation [26] and computer vision [17],as well as its recent application to improve recommendation performances [8, 31]. Chen et al. [8] designed a multimedia recommendation model which makes use of component-level and item-level attention networks to obtain different item features. Seo et al. [31] applied a local attention network to obtain user preferences or item properties, and employed the global attention network to learn global semantic meaning of review text for the review rating predictions.

Memory networks, which have been identified as effective in the field of question answering and dialogue [33], composed of two parts: an external memory and a controller. The external memory is a matrix that can increase the model's ability to store knowledge independently and save scene knowledge, and can trace long-term dependencies for inference. The controller is able to update the memory through the actions of reading, writing and deleting. The strategy of manipulating memory is used by content-based addressing, which establishes a scoring function according to a given query and a piece of text [7, 10, 13]. In the field of recommendation, Chen et al. [10] presented a memory-augmented neural network which explicitly stores and updates a user's historical records by leveraging the external memory. Ebesu et al. [13] treated the memory component's associative addressing scheme as the closest neighbourhood model to deduce user preferences.

In the current paper, an attention-based memory component is introduced to deal with the differences of interest aspect in the user-friend context, and an attention network is applied to distinguish the influences of user overall preferences.

3 Preliminaries

3.1 Definitions

In the following, we introduce the key definitions on POI recommendation, and notations used in this paper are listed in Table 1.

Definition 1 (User-friend Context). User-friend context describes the social relationship between users and their friends. It can be modeled as a graph $G_U = \{U, E_U\}$, where U denotes the set of users, and E_U denotes the set of edges among friends.

Definition 2 (POI neighbourhood Context). POI neighbourhood context captures the geographical proximity among POIs. It can be modeled as a graph $G_V = \{V, E_V\}$, where V denotes the set of POIs and E_V denotes the set of edges between nearby POIs.

Definition 3 (Check-in). Each check-in record is a triplet $ci = (u, v, t)$, which indicates that POI v was visited by user u at the specific time t.

3.2 Problem statement

Consider a set of users $U = \{u_1, u_2, ..., u_m\}$, a set of POIs $V = \{v_1, v_2, ..., v_n\}$, and a user-POI interaction matrix $Y \in R^{m\times n}$. If user $u_i \in U$ checked-in at POI $v_j \in V$, then y_{ij} equals 1, and is 0 otherwise. Given the users' check-in records, the goal of POI recommendation is to predict a rank of unvisited POIs that are potentially of interest to each user.

Table 1 List of notation

G'_u, G'_v	The friends of u and the neighbours of POI v
M, N	The number of users and POIs
$W^\star, b^\star$	The weight matrix and bias vector
q, p	The embedding of users and POIs
Y	The user-POI interaction matrix
K	The memory matrix
D	The key matrix
d	The size of the attention unites
s	The number of neighbours
γ	The geographical correlation between two POIs
λ	The regularization hyperparameter

4 The MANC model

The motivation of our work is to take full advantage of the user-social information and the POI-neighbourhood information for POI recommendations. Figure 2 presents a brief high-level overview of our model. In this model, both users and POIs are converted into dense vector representations through embedding operations. Let $p \in \mathbb{R}^d$ denote the user embedding, $q \in \mathbb{R}^d$ denote the POI embedding, and d denote the embedding size. Our model consists of two parts: a user-friend module and a POI neighbourhood module. The user-friend module is used to differentiate the influence of interest aspect between users and their friends, and to subsequently learn the effect of users' overall preferences resulting from their friends. The POI neighbourhood module is designed for distinguishing the influence of both the geographical distance and the correlation of POI's latent features across POIs.

4.1 User-friend module

Aspect-Level Attention Network In general, a user and her friends can share the same interests in certain aspects, particularly when the user has different interests. Therefore, different friends may pay different attention to the user in certain interest aspects. Since our work is based on implicit feedback, the same interest relation between users and their friends in implicit data is not explicit. It is difficult to identify what same interest aspects they share in implicit data. Inspired by the rapid development of attention mechanisms and memory networks [7, 8], we present an attention-based memory component to model the different influences between users and their friends in terms of their interest aspects, as shown in Fig. 3. The memory matrix of the component is denoted as $K \in \mathbb{Q}^{N\times d}$, and N is the memory size. The attention-based memory component accepts two inputs, user embedding p_i and user's friends' embeddings $\{p_{(i,1)}, p_{(i,2)}, ..., p_{(i,h)}\}$. The output is vector $f_{(i,h)}$, representing the relationship between u_i and the h-th friend of user u_i.

To validly capture the influence between users and their friends in same interest aspects from the user-friend context, we employ the element-wise product to learn the joint vector of the user embedding p_i and his or her friends' embedding $p_{(i,h)}$. We then normalize the joint vector to maintain the same scale by adding the denominator, as follows:

$$z = \frac{p_i \odot p_{(i,h)}}{\|p_i\| \|p_{(i,h)}\|}, \tag{1}$$

where $\odot$ is the element-wise product in Eq. (1). The dimension of joint vector $z \in \mathbb{Q}^d$ is equal to that of p_i and $p_{(i,h)}$. We subsequently determine the gain of the aspect

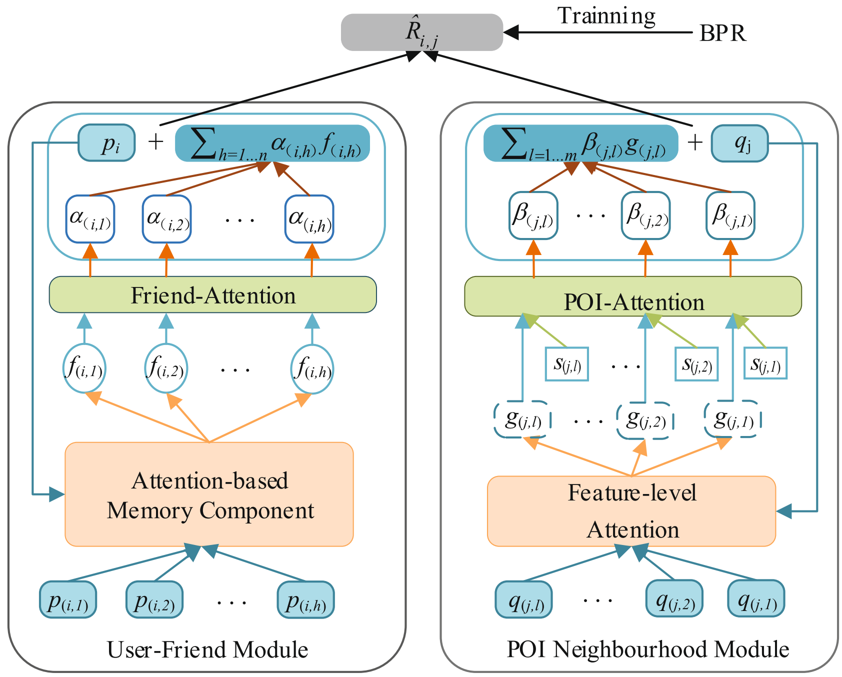


Fig. 2 The architecture of MANC that contains two main parts: a user-friend module and a POI neighbourhood module

attention score from key matrix $D \in \mathbb{Q}^{N \times d}$. Each element of aspect attention score $\delta^{\star}$ is calculated in Eq. (2) and the final aspect attention weight is obtained by normalizing $\delta^{\star}$ with the softmax function:

$$\delta_t^{\star} = z^T D_t, \tag{2}$$

$$\delta_t = \frac{exp(\delta_t^{\star})}{\sum_r exp(\delta_r^{\star})}, \tag{3}$$

where $D_t \in \mathbb{Q}^d$ and the aspect attention weight $\delta \in \mathbb{Q}^N$. We employ the memory matrix K to extend friend embedding $u_{(i,h)}$ to a matrix:

$$F_t = p_{(i,h)} \odot K_t, \tag{4}$$

where $\odot$ is the element-wise product in Eq. (4) and $F \in \mathbb{Q}^{d \times N}$ is a matrix representing the number of units that store different aspects of the preferences between a user and his or her friends. The parameter N can be considered as the number of latent aspects. Finally, the friend vector can be generated by multiplying the aspect attention weight with matrix F and performing the sum operation:

$$f_{(i,h)} = \sum_t \delta_t F_t. \tag{5}$$

Friend vector $f_{(i,h)}$ denotes the influence of user u_i's h-th friend's preferences. In addition, we can subsequently obtain user u_i's friend vectors $\{f_{(i,1)}, f_{(i,2)}, ..., f_{(i,h)}\}$ following the aspect-level attention network.

Friend-Level Attention Network As mentioned in Section 1, different friends of a user do not equally affect a user's preferences in reality. To simulate such disparate effects, we present a friend-level attention network that can automatically acquire the impact level of each friend.

The friend-level attention network aims to assign suitable weights to the user's friends. During the training of the model, when the user interacts with different POIs, the weights will change. The friend-level attention network accepts user embedding q, current item embedding p and friend vector f as the inputs, and outputs the final

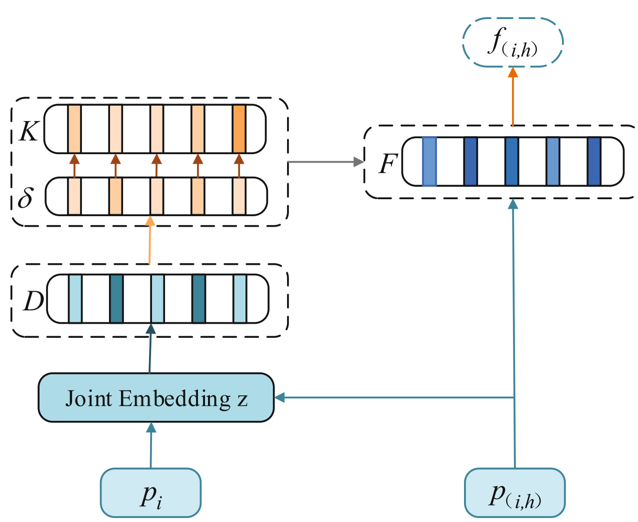


Fig. 3 The attention-based memory component

representation of user P. Through the dual layer network mechanism, we can derive a friend-level attention score defined as follows:

$$\alpha^{\star}_{(i,h)} = w_1^T \phi(W_1 p_i + W_2 q_j + W_3 f_{(i,h)} + b_1), \tag{6}$$

where $W_1 \in \mathbb{Q}^{d\times r}$, $W_2 \in \mathbb{Q}^{d\times r}$, $W_3 \in \mathbb{Q}^{d\times r}$ and $b_1 \in \mathbb{Q}^r$ are the first layer parameters, and the parameter $w_1 \in \mathbb{Q}^r$ is the second layer parameter. $\phi(x) = max(0, x)$ is the *ReLU* [28] function that was found to be better than a single layer perceptron with hyperbolic tangent nonlinearity. The final friend-level attention weight is obtained by normalizing the above attentive score using the Softmax function, which can be interpreted as the different contributions of user u_i's friends to user u_i:

$$\alpha_{(i,h)} = \frac{exp(\alpha^{\star}_{(i,h)})}{\sum_{i\in G'_{U_i}} exp(\alpha^{\star}_{(i,h)})}, \tag{7}$$

where G'_{u_i} denotes the friends of user u_i in the user-friend context. After obtaining the friend-level attention weight, the representation of user P_i is calculated follows:

$$P_i = p_i + \sum_{h\in C_i} \alpha_{(i,h)} f_{(i,h)}. \tag{8}$$

In Eq. (8), final user vector P_i contains user u_i's own preferences and the influence of his or her friends, and the addition strategy is adopted to model the preferences of users. Note that the strategy to fuse different features can be achieved through concatenation, addition and the element-wise product, all of which have been explored in many studies, including NARRE [6], A^{3}NCF [8] and SAMN [7].

4.2 POI neighbourhood module

Existing works consider the introduction of POI geographical information to effectively improve the performance of POI recommendation. In most cases, for a target POI, users are more willing to visit POIs that are closer to the target POI. Meanwhile, there is a situation where users may be attracted by the different features of POIs, which may not be explained by physical distance. To capture variations in the geographical distance and the relevance of POI features across POIs, we fuse the POI features and distance-based features to model the impact of the POI neighbourhood context. However, the features of POIs are not available in implicit data. Therefore, we apply attention networks to learn the related latent features of the POIs. The strategy of combining latent features has been widely applied for improving the efficiency of recommendations in many studies [14, 27, 34, 42].

Given POI embedding q_j and neighbourhood POIs embedding $\{q_{(j,1)}, q_{(j,2)}, ..., q_{(j,l)}\}$, we can obtain the attentive score for each POI embedding pair as follows:

$$a_{(j,l)} = q_j \odot q_{(j,l)}, \tag{9}$$

where $\odot$ is the element-wise product of Eq. (9), and $a_{(j,l)} \in \mathbb{Q}^d$ is the attentive score. After normalizing the above attentive score by using the Softmax function, the relevance vector can be generated through the follow operation:

$$g_{(j,l)} = \frac{exp(a_{(j,l)})}{\sum_d exp(a_{(j,l)})}. \tag{10}$$

The relation vector $g_{(j,l)} \in \mathbb{Q}^d$ can be interpreted as the importance of the correlation between POI q_j and q_j's l-th neighbour on the latent features. To reflect the efficacy of the geographical influence, we incorporate the geographical distance property using a Gaussian radial basis function kernel (RBF kernel) which is defined as follows:

$$S_{(j,l)} = exp(-\gamma \| L_j - L_{(j,l)} \|), \tag{11}$$

where L_j and $L_{(j,l)}$ denote the geographical coordinates of two POIs. The RBF kernel is a distance decay function with value range limited to [0,1]. To simplify the calculations, the value of $S_{(j,l)}$ is set as 0 if it is less than 0.1. The hyper-parameter $\gamma > 0$ controls the geographical correlation between two POIs, and a large γ will generate a large $S_{(j,l)}$. In our work, we calculate the pair-wise RBF value of each pair in advance to obtain the distance-based features, and then we fuse the distance-based features and the related latent features of the POIs to generate the POI-level attention score of the POI neighbourhood influence:

$$\beta^{\star}_{(j,l)} = w_2^T \phi(W_4 p_i + W_5 q_j + W_6 g_{(j,l)} + W_7 S_{(j,l)} + b_2), \tag{12}$$

where $W_4 \in \mathbb{Q}^{d\times r}$, $W_5 \in \mathbb{Q}^{d\times r}$, $W_6 \in \mathbb{Q}^{d\times r}$, $W_7 \in \mathbb{Q}^{d\times r}$, and $b_2 \in \mathbb{Q}^r$, $w_2 \in \mathbb{Q}^r$ are the parameters of our model, and $\phi(x)$ is the *ReLU* activation function. To obtain the final POI-level attention weights, we use the Softmax function to normalize the above attentive score as in Eq. (13). This practice is often applied in neural attention network.

$$\beta_{(j,l)} = \frac{exp(\beta^{\star}_{(j,l)})}{\sum_{j\in G'_{V_j}} exp(\beta^{\star}_{(j,l)})}, \tag{13}$$

where G'_{v_j} denotes the set of POI q_j's neighbours. After we achieve the POI-level attention weight for each POI, the final representation of POI Q_j is calculated as follows:

$$Q_j = q_j + \sum_{l \in G_j} \beta_{(j,l)} g_{(j,l)}. \tag{14}$$

Similar to the generation of the user vector from the user-friend module, the final POI vector considers both the POI's own features and the influence of the neighbourhood containing geographical distance and the relevance of the POI's latent features.

4.3 Prediction

The prediction part of our model is based on Matrix Factorization [16] which effectively models the implicit feedback and is widely applied in many studies [14, 30, 35, 42]. After we obtain the optimized user vectors and POI vectors, the recommendation task can be reformulated as a ranking task that is based on the estimated score for all the POIs within the dataset. The prediction score can be calculated as follows:

$$\hat{R}_{ij} = \left(p_i + \sum_{h \in C_i} \alpha_{(i,h)} f_{(i,h)}\right)^T \left(q_j + \sum_{l \in G_j} \beta_{(j,l)} g_{(j,l)}\right). \tag{15}$$

4.4 Model training

The recommendation process is a personalized ranking task and our goal is to study parameters of MANC with a ranking objective. In order to ensure that visited POIs are ranked higher than those that are unvisited, we apply a Bayesian Personalized Ranking (BPR) objective function as follows:

$$\mathcal{L}_{BPR} = \sum_{(i,j,k) \in \mathbb{S}} -ln\sigma\left(\hat{R}_{ij} - \hat{R}_{ik}\right) + \lambda_\theta\left(\|\Theta\|^2\right), \tag{16}$$

where σ is the logistic sigmoid function, λ_θ is specific regularization hyperparameter to prevent overfitting, and $\mathbb{S}$ denotes the set of training instances $\mathbb{S} := \{(i, j, k) | j \in \mathbb{S}^+ \wedge k \notin \mathbb{S}^+\}$, where $\mathbb{S}^+$denotes the set of POIs that has been visited by user u_i.

This objective function is intended to maximize the difference between visited and unvisited POIs. We apply small-batch Adagrad [12] as the optimizer to optimize the parameters of MANC. In particular, Adagrad can adaptively adjust the learning rate in the training stage, alleviates the pain of selecting the appropriate learning rate, and accelerates convergence compared to the traditional stochastic gradient descent. The mini-batch training algorithm is shown in Algorithm 1.

Algorithm 1 Training algorithm.

Input: Users U, user friends G'_U, POIs V, POI neighbours G'_V, geographical distance S;
Output: Recommend top-K POIs;
Initialize all parameters;
numBatches = M $/batchSize$;
while $iter < numIteration$ **do**
  Shuffle(U, G'_u, V, G'_v)
  for $batchID = 0$; $batchID < numBatches$; $batchID{+}{+}$ **do**
    $Ubatch, G'_U batch, Vbatch, G'_V batch$ =ExtractBatchData ($batchID, U, G'_U, V, G'_V$);
    Apply Eqs.(1)~(8) to get P_i;
    Apply Eqs.(9)~(14) to get Q_j;
    Apply Eq. (15) to calculate predicted probability $\hat{R}_{ij}$;
    Apply Eq. (16) to obtain back-propagate the error and update parameters through the entire network;
  end
end

5 Experiments

In this section, we first introduce the experimental settings in detail, and then empirically evaluate the performance of our model on two real-world datasets.

5.1 Experimental settings

Dataset We utilize two real-world datasets to evaluate the performance of our model, which are Gowalla and Yelp. The Gowalla dataset with rich geographic information and social information was generated from February 2009 to October 2010. The Yelp dataset contains a large number of geotagged businesses (considered POIs), reviews and social connections among users. Both datasets are preprocessed to ensure users with fewer than 15 check-in POIs and POIs with less than 10 visited users. The Gowalla dataset contains 3534 users, 5349 POIs and 298833 check-ins. There are 3118 users, 2860 POIs and 109187 check-ins in the Yelp dataset. For each dataset, we partition the earliest 80% of check-ins as the training data and employ the remaining as testing data.

Evaluation Metrics To evaluate the performance of all models, we use four metrics, i.e., Precision@K, Recall@K, NDCG@K (Normalized Discounted Cumulative Gain) and MAR@K (Mean Average Precision). Each metric reflects the evaluation results of different aspects: the precision and recall consider the number of correct recommendation results in the top-k recommendation list while the NDCG and MAP focus on the rank position of the correct recommendation in the top-k recommendation list. The evaluation metrics are formally defined as follows:

$$Precision@k = \frac{1}{M}\sum_{i=1}^{M}\frac{S_i(k)\cap T_i}{k}, \tag{17}$$

$$Recall@k = \frac{1}{M}\sum_{i=1}^{M}\frac{S_i(k)\cap T_i}{|T_i|}, \tag{18}$$

$$MAP@k = \frac{1}{M}\sum_{i=1}^{M}\frac{\sum_{j=1}^{k} p(j)\times rel(j)}{|T_i|}, \tag{19}$$

$$DCG_k = \sum_{j=1}^{k}\frac{2^{rel(j)}-1}{log_2(j+1)}, \tag{20}$$

$$NDCG@k = \frac{1}{M}\sum_{i=1}^{M}\frac{DCG_k}{IDCG_k}, \tag{21}$$

where $S_i(k)$ denotes the recommendation list of the top-k POIs that have been unvisited by user u_i during testing. T_i denotes the list of POIs that are visited by user u_i during testing. $P(j)$ denotes the accuracy of a rank list that is cut from 1 to j. $rel(j)$ equals 1 if the user visited the location in the testing, otherwise, $rel(j) = 0$. The NDCG is determined by the DCG (Discounted Cumulative Gain) and the IDCG (Ideal Discounted Cumulative Gain), and the IDCG is the best ranking function value for DCG.

Comparison Methods To estimate the performance of our proposed model, we experimentally compare MANC with the following baseline methods.

BPRMF. [30] The Bayesian personalized ranking optimizes the ranking order of observed and unobserved locations using a pairwise loss. It is a highly competitive implicit feedback recommendation method.

LORE. [41] This is a sequential-based method that mines sequential patterns from location sequences and incorporates an Additive Markov Chain to predict the probability of a user visiting a POI.

MGMPFM. [11] This method applies multiple Gaussian distributions to simulate the regions of activity for each user and combines matrix factorization for POI recommendation.

PACE. [35] This method utilizes a deep neural model for POI recommendation. It trains the embedding of users and POIs by combining MLP and the various context associated with users and POIs.

SAE-NAD. [27] This is a recently proposed autoencoder-based model that adopts a self-attention encoder to learn the users' degrees of preference in multiple aspects and applies a neighbor-aware decoder to incorporate the geographical context information.

BGMF. The method is a variant of GMF (generalized matrix factorization) [14] and only learns the user-POI interactions based on user embeddings and POI embeddings with a pairwise loss that is as same as MANC.

Parameter Settings When comparing MANC with baselines, the parameters are set as follows during the experiment. For both datasets, we set the batch size to 256, the learning rate $lr = 0.05$, the number of neighbourhoods k to 200, the geographical correlation level γ to 60 and the embedding size to 256. For the Gowalla dataset, the number of memory slices N is set to 32, the regularization parameters $\lambda = 0.002$ and the size of the attention unites $d = 64$. For the Yelp dataset, the number of memory slices N is set to 16, the regularization parameter $\lambda = 0.005$ and the size of the attention unites $d = 32$. In addition to these default values, we decide to test different parameter values to observe how they affect the performance of MANC in Section 5.4. For all baselines, the parameters were initialized as in the corresponding papers. Experiments were performed using the Gowalla and Yelp datasets, with the careful tuning of the parameters to achieve optimal performance.

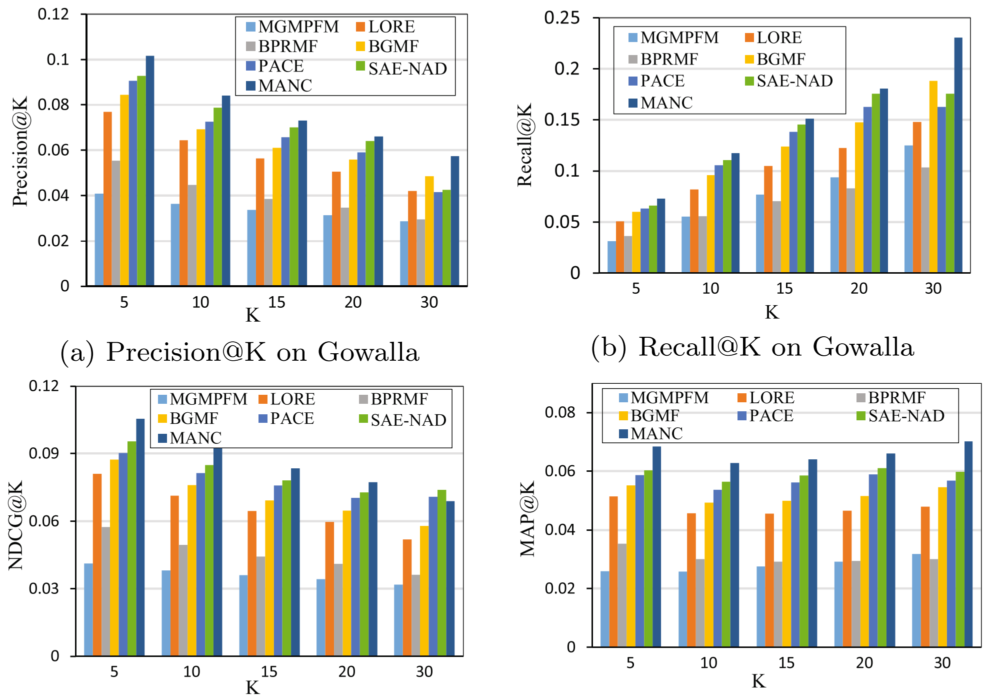
5.2 Overall performance

Figures 4 and 5 show the prediction results of our model and the baseline methods for the four metrics on the two datasets.

Combined with the experimental results, we can get the following conclusions about our model. First, the scores of the Gowalla dataset are generally higher than those of the Yelp dataset. This is due to the user's richer check-in history in previous datasets of the Gowalla dataset. Second, our MANC model achieves remarkable results on the four metrics, which illustrates the superiority of our model. Compared with the strongest baseline SAE-NAD, our model relatively improves the precision by approximately 11.7%, the recall by approximately 6.2%, the NDCG by approximately 4.1% and the MAP by approximately 12% on the Gowalla dataset. The performance improvement on the Yelp dataset is 8.5% on the precision, 7.7% on the recall, 1.7% on the NDCG and 4.8% on the MAP. The results show the benefits of exploiting context attention that can increase the accuracy of recommendations. Third, SAE-NAD performs better than all other baseline methods since it applies a deep neural structure to model the user preferences in the check-in data and captures the geographical influence between checked-in POIs and unvisited POIs.

In addition, we can also obtain other observations. First, both PACE and BGMF with nonlinear activation function

Fig. 4 The performance comparison on Gowalla



(a) Precision@K on Gowalla

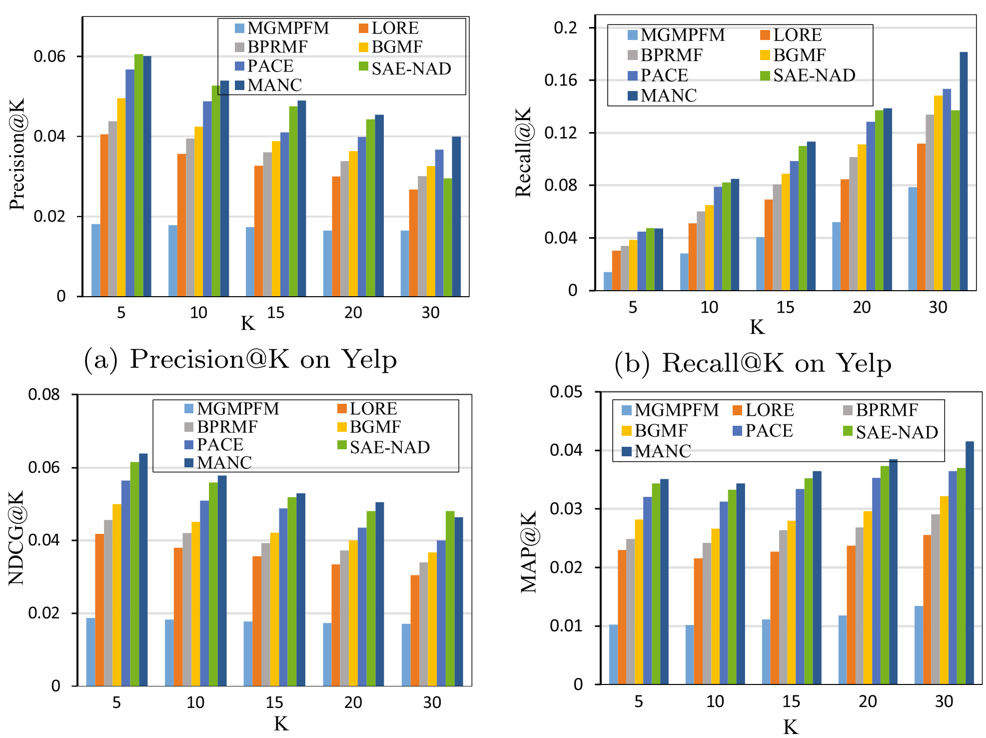
(b) Recall@K on Gowalla

(c) NDCG@K on Gowalla

(d) MAP@K on Gowalla

model user-POI interactions for implicit feedback, but the performance of PACE is better than that of BGMF. One possible reason is that PACE applies contextual information to model the important influence of the user-POI interactions while BGFM does not consider other contextual information in the check-in data. Second, BPRMF is superior to LORE and MGMPFM. The reason is that BPRMF is designed for implicit feedback and leverages pair loss

Fig. 5 The performance comparison on Yelp

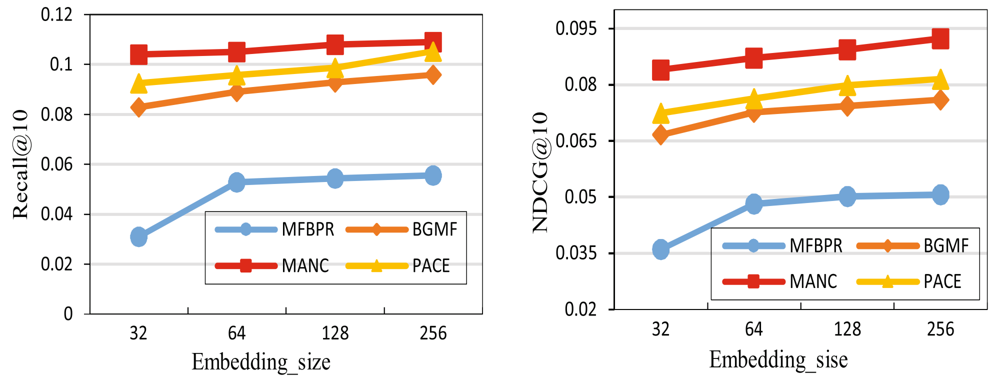


(a) Precision@K on Yelp

(b) Recall@K on Yelp

(c) NDCG@K on Yelp

(d) MAP@K on Yelp



(a) Recall@10 on Gowalla (b) NDCG@10 on Gowalla

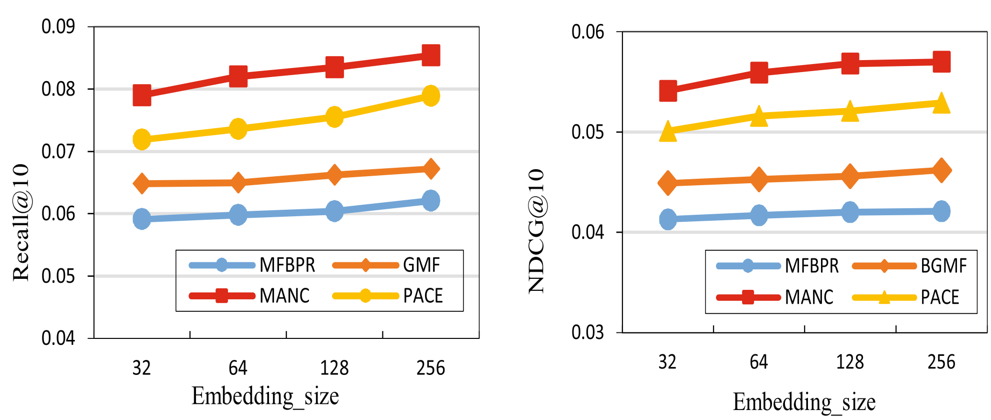
Fig. 6 The performance comparison on Gowalla

function to train and optimize the correct ranking for POI recommendation. Third, the performance of MGMPFM is the worst since it is not good at modelling the preferences of users from implicit feedback, but rather it just directly models users' check-in frequency. LORE based on the probabilistic factor model, which captures users' check-in behaviours from a check-in POI stream and employs an Additive Markov Chain to enhance the sequential influence.

Meanwhile, we also implement experiments to examine the effect of the embedding size. Since our method is embedded in the same way as MFBPR, BGMF, and PACE, we chose to compare MANC with these three methods. Figures 6 and 7 show the performance of Recall@10 and NDCG@10 for the embedding size. As seen from these figures, the performance of our model is better than MFBPR, BGMF, and PACE on both datasets. Furthermore, a larger embedding size can increase the performance of all models on two datasets. The results show that a larger dimension can accommodate more latent factors of users and POIs and improve the modelling capability.

5.3 Impacts of context attention networks

To demonstrate the utility of the different context-based attention modules embed-ded in the model, we solely compare the performance of each component. Here, we denote the user-friend module with BGMF as MANC-U and the POI neighbourhood module with BGMF as MANC-P. The individual components of the proposed model are shown in Table 2. There are two observations from the table. (1) MANC generally outperforms MANC-P and MANC-U on both datasets. The reason that MANC performs better is that MANC can capture the influence of both the user-friend context and POI neighbourhood context. (2) MANC-U and MANC-P beat BGMF in all cases. For example, for the top-10 POI recommendation, MANC-U beats BGMF by 20.8% and MANC-P beats BGMF by 19.2% w.r.t. Precision@10 on Gowalla, showing the effectiveness of applying the multi-attention network on modelling the contextual information for users and POIs.



(a) Recall@10 on Yelp (b) NDCG@10 on Yelp

Fig. 7 The performance comparison on Yelp

Table 2 The performance of the context-based attention module of MANC on Gowalla and Yelp

Gowalla	Precision@10	Recall@10	NDCG@10	MAP@10
BGMF	0.06929	0.09592	0.07605	0.04936
MANC-U	0.08372	0.11698	0.09168	0.0618
MANC-P	0.08265	0.11555	0.0905	0.06145
MANC	0.08412	0.1175	0.09231	0.06276
Yelp	Precision@10	Recall@10	NDCG@10	MAP@10
BGMF	0.04249	0.06521	0.04506	0.02667
MANC-U	0.05365	0.08364	0.05591	0.03401
MANC-P	0.05336	0.08312	0.05594	0.03383
MANC	0.05439	0.08517	0.05709	0.03404

5.4 Sensitivity of parameters

We evaluate the performance of MANC for different parameters: the regularization parameter λ, the size of the attention unites d and the number of neighbours s in the POI neighbourhood context. In the process of evaluating MANC, it needs more time to calculate the ranking of all POIs for each user many times. Therefore, we evaluate the performance by using the leave-one-out scheme that is widely used in many literatures [8, 14, 35, 42]. We implement the strategy that selects 100 POIs that a user has not visited and the 100 POIs are mixed with the user's groundtruth POIs. Then, we rank the 100 POIs along with the test POI and calculate the HR@10 (Hit Ratio) that determines whether the test POI is ranked in the top-10 list. The results for the HR@10 are reported for both datasets. We can make the following observations from Table 3.

(1) Fewer neighbours s decreases the performance of MANC; and as d increases, the HR@10 becomes steady. (2) The HR@10 becomes large when the size of the attention unites d increases, but the influence on the performance of MANC is marginal. (3) MANC achieves the best performance with $\lambda = 0.002$ on Gowalla and $\lambda = 0.004$ on Yelp, which indicates that the number of neighbours should be set neither too small nor too large.

Table 3 HR@10 w.r.t different parameters in MANC

HR@10 w.r.t s						
Dataset	s=50	s=100	s=150	s=200	s=300	
Gowalla	0.756	0.764	0.771	0.783	**0.790**	
Yelp	0.643	0.655	0.651	**0.659**	0.656	
HR@10 w.r.t d						
Dataset	d=8	d=16	d=32	d=64		
Gowalla	0.768	0.773	0.784	**0.786**		
Yelp	0.646	0.650	0.656	**0.658**		
HR@10 w.r.t λ						
Dataset	λ=0	λ=0.001	λ=0.002	λ=0.003	λ=0.004	λ=0.005
Gowalla	0.711	0.782	**0.791**	0.781	0.763	0.752
Yelp	0.575	0.631	0.640	0.656	**0.661**	0.649

6 Conclusion

The MANC model presented in this paper provides an approach for making full use of the contextual information of both users and POIs for POI recommendation. In the user-friend module of MANC, the aspect-level differences among user-friend common preferences can be calculated by an attention-based memory, and the different influences of user's different friends can be adaptively discriminated by assigning importance scores. In the POI neighbourhood module of MANC, both the influence caused by geographical distance and the influence caused by latent features of POIs can be taken into account. Experimental results show that our approach works well and outperforms state-of-the-art methods.

Future research will focus on extending MANC model to deal with additional auxiliary information such as user reviews, POI tags and temporal signals. Further to this, we aim to construct a POI recommendation system with knowledge graphs which contain rich semantic information of users and POIs [5, 39]. With the help of knowledge graphs, we believe that both the performance of recommendation and the interpretability of recommendation results could be improved.

Acknowledgements This work is supported by the Natural Science Foundation of China (Nos. U1811264,U1711263,61966009), the Natural Science Foundation of Guangxi Province (No. 2018GXNSFDA281045, No. 2020GXNSFAA159055), and the Guangxi Innovation-Driven Development Project (No. AA17202024).

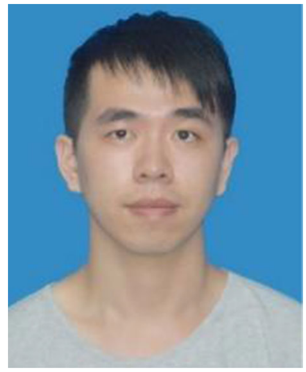
References

1. Ahmadian S, Joorabloo N, Jalili M, Meghdadi M, Afsharchi M, Ren Y (2018) A temporal clustering approach for social recommender systems. In: Proceedings of the 10th international conference on advances in social networks analysis and mining, IEEE Computer Society, pp 1139–1144
2. Ahmadian S, Afsharchi M, Meghdadi M (2019) An effective social recommendation method based on user reputation model and rating profile enhancement. J Inf Sci 45(5):607–642
3. Ahmadian S, Joorabloo N, Jalili M, Ren Y, Meghdadi M, Afsharchi M (2020) A social recommender system based on reliable implicit relationships. Knowl Based Syst 192:105,371
4. Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. J Mach Learn Res 3(2):1137–1155
5. Chang L, Shi Z, Gu T, Zhao L (2012) A family of dynamic description logics for representing and reasoning about actions. J Autom Reasoning 49(1):1–52
6. Chen C, Zhang M, Liu Y, Ma S (2018) Neural attentional rating regression with review-level explanations. In: Proceedings of the 2018 world wide Web conference international world wide Web conferences steering committee, pp 1583–1592

7. Chen C, Zhang M, Liu Y, Ma S (2019) Social attentional memory network: Modeling aspect-and friend-level differences in recommendation. In: Proceedings of the 12th ACM international conference on Web search and data mining, ACM, pp 177–185
8. Chen J, Zhang H, He X, Nie L, Liu W, Chua TS (2017) Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, ACM, pp 335–344
9. Chen Q, Xu J, Koltun V (2017) Fast image processing with fully-convolutional networks. In: Proceedings 2017 IEEE international conference on computer vision, IEEE Computer Society, pp 2516–2525
10. Chen X, Xu H, Zhang Y, Tang J, Cao Y, Qin Z, Zha H (2018) Sequential recommendation with user memory networks. In: Proceedings of the 11th ACM international conference on Web search and data mining, ACM, pp 108–116
11. Cheng C, Yang H, King I, Lyu MR (2012) Fused matrix factorization with geographical and social influence in location-based social networks. In: Proceedings of the 26th AAAI conference on artificial intelligence, pp 17–23
12. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 12(7):2121–2159
13. Ebesu T, Shen B, Fang Y (2018) Collaborative memory network for recommendation systems. In: Proceedings of the the 41st international ACM SIGIR conference on research and development in information retrieval, ACM, pp 515–524
14. He X, Liao L, Zhang H, Nie L, Hu X, Chua TS (2017) Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide Web, International World Wide Web Conferences Steering Committee, pp 173–182
15. Hinton G, Deng L, Yu D, Dahl G, Ar Mohamed, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Kingsbury B et al (2012) Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29(6):82–97
16. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. IEEE Computer 42(8):30–37
17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Proceedings of the 26th annual conference on neural information processing systems, pp 1097–1105
18. Li H, Ge Y, Hong R, Zhu H (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, pp 975–984
19. Li X, Cong G, Li XL, Pham TAN, Krishnaswamy S (2015) Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, ACM, pp 433–4421
20. Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y (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, pp 831–840
21. Liu B, Fu Y, Yao Z, Xiong H (2013) Learning geographical preferences for point-of-interest recommendation. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 1043–1051
22. Liu B, Xiong H, Papadimitriou S, Fu Y, Yao Z (2014) A general geographical probabilistic factor model for point of interest recommendation. IEEE Trans Knowl Data Eng 27(5):1167–1179
23. Liu B, Su Y, Zha D, Gao N, Xiang J (2019) Carec: Content-aware point-of-interest recommendation via adaptive bayesian personalized ranking. Australian Journal of Intelligent Information Processing Systems 15(3):61–68
24. Liu Y, Wei W, Sun A, Miao C (2014) Exploiting geographical neighborhood characteristics for location recommendation. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management, ACM, pp 739–748
25. Liu Y, Pham TAN, Cong G, Yuan Q (2017) An experimental evaluation of point-of-interest recommendation in location-based social networks. Proceedings of the VLDB Endowment 10(10):1010–1021
26. Luong T, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 conference on empirical methods in natural language processing, the association for computational linguistics, pp 1412–1421
27. Ma C, Zhang Y, Wang Q, Liu X (2018) Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence. In: Proceedings of the 27th ACM international conference on information and knowledge management, ACM, pp 697–706
28. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning, Omnipress, pp 807–814
29. Rahmani HA, Aliannejadi M, Ahmadian S, Baratchi M, Afsharchi M, Crestani F (2019) Lglmf: Local geographical based logistic matrix factorization model for POI recommendation. In: Proceedings of the 15th Asia information retrieval societies conference, Springer, pp 66–78
30. Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2009) Bpr: Bayesian personalized ranking from implicit feedback. In: Proceedings of the 25th conference on uncertainty in artificial intelligence, AUAI Press, pp 452–461
31. Seo S, Huang J, Yang H, Liu Y (2017) Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In: Proceedings of the 11th ACM conference on recommender systems, ACM, pp 297–305
32. Wang H, Shen H, Ouyang W, Cheng X (2018) Exploiting poi-specific geographical influence for point-of-interest recommendation. In: Proceedings of the 27th international joint conference on artificial intelligence, pp 3877–3883
33. Weston J, Chopra S, Bordes A (2015) Memory networks. In: Proceedings of the 3rd international conference on learning representations
34. Xie M, Yin H, Wang H, Xu F, Chen W, Wang S (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, pp 15–24
35. Yang C, Bai L, Zhang C, Yuan Q, Han J (2017) Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 1245–1254
36. Ye M, Yin P, Lee WC (2010) Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, ACM, pp 458–461

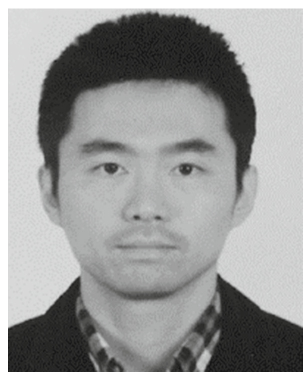
37. Ye M, Yin P, Lee WC, Lee DL (2011) Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval, ACM, pp 325–334
38. Yin H, Zhou X, Shao Y, Wang H, Sadiq S (2015) Joint modeling of user check-in behaviors for point-of-interest recommendation. In: Proceedings of the 24th ACM international on conference on information and knowledge management, ACM, pp 1631–1640
39. Yochum P, Chang L, Gu T, Zhu M (2020) Linked open data in location-based recommendation system on tourism domain: a survey. IEEE Access 8:16,409–16,439
40. Zhang JD, Chow CY (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, pp 443–452
41. Zhang JD, Chow CY, Li Y (2014) Lore: Exploiting sequential influence for location recommendations. In: Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information Systems, ACM, pp 103–112
42. Zhang S, Cheng H (2018) Exploiting context graph attention for poi recommendation in location-based social networks. In: Proceedings of the 23rd international conference on database systems for advanced applications, Springer, pp 83–99
43. Zhang Z, Liu Y, Zhang Z, Shen B (2019) Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation. World Wide Web 22(3):1135–1150
44. Zhao S, Zhao T, Yang H, Lyu MR, King I (2016) Stellar: spatial-temporal latent ranking for successive point-of-interest recommendation. In: Proceedings of the 30th AAAI conference on artificial intelligence, AAAI Press, pp 315–322
45. Zhao S, Zhao T, King I, Lyu MR (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, pp 153–162
46. Zheng L, Noroozi V, Yu PS (2017) Joint deep modeling of users and items using reviews for recommendation. In: Proceedings of the 10th ACM international conference on Web search and data mining, ACM, pp 425–434

Liang Chang received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, in 2008. He is currently a Professor with the School of Computer Science and Information Security, Guilin University of Electronic Technology, China. His research interests include data and knowledge engineering, formal methods, and intelligent planning.



Wei Chen is currently pursuing the Ph.D. degree with the School of Computer and Information Technology, Beijing Jiaotong University, China. His main research interests include knowledge graph, recommendation system.

Jianbo Huang is currently working towards the Master degree at the School of Computer Science and Information Security, Guilin University of Electronic Technology, China. His research interests include privacy preservation, recommendation system.



Chenzhong Bin received the B.Eng. and M.Eng. degrees in computer science and technology from North West Polytechnical University, China, in 2002 and 2005. He is currently pursuing the Ph.D. degree with the School of Information and Communication, Guilin University of Electronic Technology. His main research interests include data mining, machine learning and IOT applications.

Wenkai Wang is a graduate student at the School of Computer Science and Information Security, Guilin University of Electronic Technology, China. His research interest include question answering system, natural language processing.

Affiliations

Liang Chang[1,2] · **Wei Chen**[3] · **Jianbo Huang**[1] · **Chenzhong Bin**[1] · **Wenkai Wang**[1]

Liang Chang
changl@guet.edu.cn

Jianbo Huang
hollisterwong@gmail.com

Chenzhong Bin
binchenzhong@163.com

Wenkai Wang
ayuwwk@foxmail.com

1 Guangxi Key lab of Trusted Software, Guilin University of Electronic Technology, Guilin, China

2 Center of Guangxi Province for Collaborative Innovation in the Technology of IOT and the Industrialization, Guilin University of Electronic Technology, Guilin, China

3 Beijing Jiaotong University, Beijing, China