Exploring the Context of Locations for Personalized Location Recommendations

Xin Liu, Yong Liu and **Xiaoli Li**
Institute for Infocomm Research (I^2R), A*STAR, Singapore
{liu-x, liuyo, xlli}@i2r.a-star.edu.sg

Abstract

Conventional location recommendation models rely on users' visit history, geographical influence, temporal influence, etc., to infer users' preferences for locations. However, systematically modeling a location's *context* (i.e., the set of locations visited before or after this location) is relatively unexplored. In this paper, by leveraging the Skip-gram model, we learn the latent representation for a location to capture the influence of its context. A pair-wise ranking loss that considers the confidences of observed user preferences for locations is then proposed to learn users' latent representations for personalized top-*N* location recommendations. Moreover, we also extend our model by taking into account temporal influence. Stochastic gradient descent based optimization algorithms are developed to fit the models. We conduct comprehensive experiments over four real datasets. Experimental results demonstrate that our approach significantly outperforms the state-of-the-art location recommendation methods.

1 Introduction

The majority of location recommendation models are built on top of collaborative filtering techniques by focusing on different aspects such as geographical influence [Ye *et al.*, 2011; Cheng *et al.*, 2012; Liu *et al.*, 2013a], temporal influence [Gao *et al.*, 2013; Yuan *et al.*, 2013; 2014], and semantic influence [Liu and Xiong, 2013; Hu and Ester, 2013; Liu *et al.*, 2013b]. Although explicitly modeling such influences indeed evidently improves the quality of location recommendations, most existing models directly predict a user's preference for a location without deeply investigating the context of the location, which is defined as the set of locations that are immediately visited before or after the location. Failing to handle this issue makes the recommendations not only inaccurate, but also less interpretable.

Existing models that consider the "relationship" of consecutively visited locations mainly rely on Markov chain (i.e., only latest location is considered when modeling the location sequence) [Cheng *et al.*, 2013; Chen *et al.*, 2011]. However, in reality, a user's visit at a location may be influenced or reflected by not only a set of her previously visited locations, but also the locations visited by her after the target location. In this paper, we endeavor to explore the context of locations to better understand users' visit behaviors. To this end, we learn the latent representations of locations by leveraging the Skip-gram model [Mikolov *et al.*, 2013a; 2013b; 2013c], which has been widely applied to various Natural Language Processing (NLP) tasks.

The main reason for choosing the Skip-gram model is that traditional recommendation approaches (e.g., collaborative filtering) cannot capture the context of a location. However, if we treat a user's consecutively visited locations as a trajectory that reflects her visit patterns, this is analogous to that she writes a sentence to express her semantic meanings. Thus, it is possible to leverage NLP methods to model users' mobility patterns. Moreover, compared to other NLP methods like topic modeling, Skip-gram model is capable of accurately modeling the context (i.e., surrounding words) of the target word.

To be specific, our approach first sorts each user's visits in chronological order. We treat each location as a word and each user's visited locations as a sentence, so that we have a document, represented as a collection of sentences, denotes all users' visited locations. The collection of all unique locations serve as the location vocabulary. Then, we apply the Skip-gram model to learn the latent presentation of each word (i.e., each location). Essentially, the latent representations of locations are learned by incorporating the influence of each location's context (i.e., a set of locations that were visited immediately before or after the location). Since the visited locations of each user are organized in the order of visiting time, the geographical influence between locations has been implicitly encoded. Note that although the locations visited after the target location are unknown when making recommendations in practice, such information is still utilized in training the Skip-gram model to learn high quality latent representations of locations.

By leveraging the latent representations of locations, we can easily find similar locations to make recommendations heuristically. However, such representations are learned from all users' visit history (i.e., global pattern), thus suffering from lack of personalization. To support personalized recommendations, we exploit latent factor model to learn the latent representation of each user to predict her personalized preferences for locations. A latent factor vector is assigned to every user, where the users are forced to share the same latent space with locations. In the literature, the Weighted Approximately Ranked Pairwise (WARP) loss [Weston *et al.*, 2010] is one of the most effective loss functions developed to learn latent factor models for top-*N* recommendation tasks. Practically, a user may visit a location multiple times, thus the count of visit is an important factor to measure users' preferences. To capture this factor, we propose the C-WARP loss, which introduces the confidence of preference (measured by visit count) for learning users' latent representations. The experimental results also indicate that the location recommendation accuracy can be improved by the proposed C-WARP loss.

In order to enable temporal-aware location recommendations, we propose a time-aware extension based on the C-WARP loss. Specifically, when measuring a user's time-aware preference for a location, besides the user's direct preference for the location, we also consider the user's preference for the time frame of visiting, as well as the correlation between the corresponding location and time frame. A latent factor vector is assigned to each time frame, which shares the same latent space with users and locations. We develop Stochastic Gradient Descent (SGD) optimization procedures to fit the models. In this paper, although the context of locations is modeled, we only focus on one-step location recommendations, sequential location recommendation task [Cheng *et al.*, 2013] is treated as the immediate future work.

2 Related Work

In [Ye *et al.*, 2011],the geographical influence was captured by assuming that users' geographical movements follow power-law distribution. A fused framework was proposed to incorporate users' preferences, geographical and social influences into one recommendation process. Cheng et al. investigated users' multi-center check-in behaviors and proposed a Multi-center Gaussian Model (MGM) [Cheng *et al.*, 2012]. A matrix factorization model was used to combine users' preferences, social influence, and geographical influence for recommendation. In [Liu *et al.*, 2013a], a geographical probabilistic factor analysis framework was proposed to model geographical influence. Users' preferences were modeled by treating the check-in count data as users' implicit feedback.

Recent research started exploring the temporal influence on location recommendations. Yuan et al. proposed a time-aware model to recommend locations at a specified time in a day [Yuan *et al.*, 2013]. A user-based collaborative filtering model was used to incorporate temporal information. On top of matrix factorization, Gao et al. proposed a novel location recommendation model by modeling the consecutiveness and non-uniformness of users' daily check-in preferences [Gao *et al.*, 2013].

Another line of research was to leverage the semantic information such as location categories, tags, tips, and reviews [Kurashima *et al.*, 2013; Liu and Xiong, 2013; Hu and Ester, 2013]. In [Liu and Xiong, 2013], Liu and Xiong proposed a topic and location aware recommendation model. Latent Dirichlet Allocation (LDA) was applied to infer users' interest topics by mining textual contents that were associated with locations. A Topic and Location-aware Probabilistic Matrix Factorization (TL-PMF) model was developed for location recommendations by matching the user interests to the location topics. Yang et al. proposed a hybrid user location preference model by combining the preferences inferred from users' check-in behaviors and textual contents (e.g., tips) using sentiment analysis techniques [Yang *et al.*, 2013].

Modelling the context of locations for location recommendations was relatively less explored, and the existing solutions mainly relied on Markov models. In [Mathew *et al.*, 2012], users' check-in histories were clustered, and for each cluster, a Hidden Markov Model (HMM) was built, treating location characteristics as unobservable parameters. In [Sang *et al.*, 2012], the authors proposed a probabilistic approach to model a sequence of locations and the location categories, based on which, the system was able to recommend consecutive activities and locations on the move. The transition probability from one location to another was derived via a Markov chain, considering the context and historical visit behaviors. Similar methods included [Zhang *et al.*, 2014; Cheng *et al.*, 2013; Chen *et al.*, 2011; Gambs *et al.*, 2012], which employed different variants of Markov models.

3 Our Approach

This section elaborates our context-aware location recommendation model. In Section 3.1, we learn latent representations of locations using the Skip-gram model. In Section 3.2, by minimizing the C-WARP loss, a personalized top-*N* recommendation method is presented. A time-aware extension is introduced in Section 3.3.

3.1 Learning Location Representations

We denote the set of users by $\mathbf{U} = \{u_1, u_2, ...\}$ and the set of locations by $\mathbf{L} = \{l_1, l_2, ...\}$. For each user u, her historical visit records (in chronological order) is denoted by $\mathbf{C}_u = \{c_1, c_2, ...\}$, where $c_i =< l_i, t_i >$ is the i^{th} visit of u, consisting of location l_i and the corresponding temporal information t_i. Note that l_i contains the descriptive information such as longitude and latitude. Moreover, we build a location corpus as the input of the Skip-gram model. Here, each location corresponds to a word, and accordingly, for each user, her visited locations correspond to a sentence. By aggregating all users' historical visits, a location corpus is constructed.

For every location $l \in \mathbf{L}_u$, we find its context $C(l)$, which is defined as the locations that are visited before or after l within a predefined window size (see Fig. 1 as an example). The objective of the Skip-gram model is to maximize the following location corpus probability:

$$\arg\max \prod_{l \in \mathbf{L}} [\prod_{l_c \in C(l)} p(l_c|l)], \tag{1}$$

where $p(l_c|l)$ is estimated using the softmax function: $p(l_c|l) = \frac{\exp(\mathbf{v}_c^\top \mathbf{v}_l)}{\sum_{x \in \mathbf{L}} \exp(\mathbf{v}_x^\top \mathbf{v}_l)}$, where $\mathbf{v}_l \in \mathbb{R}^{D \times 1}$ and $\mathbf{v}_c \in \mathbb{R}^{D \times 1}$ are the latent representations of the target location l and the corresponding context location l_c, respectively. D is the dimensionality of the latent space.

Typically, the size of $\mathbf{L}$ is very large, thus directly optimizing Eq. (1) is usually infeasible. In this work, we adopt

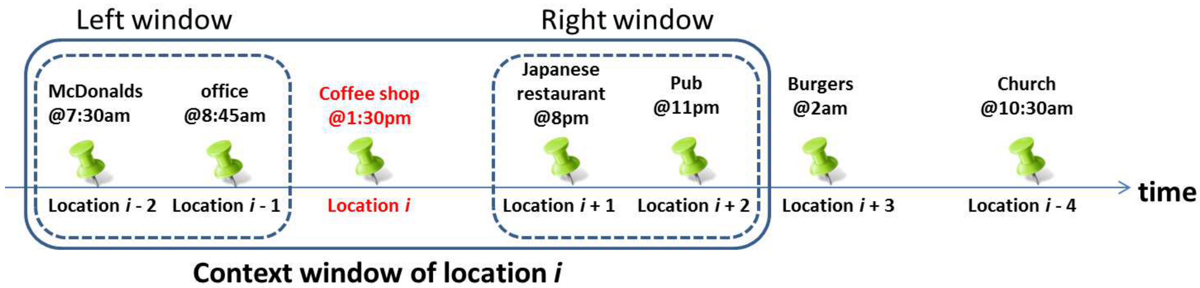


Figure 1: An example of the context window of a user's i^{th} visited location (coffee shop). The window includes 4 locations visited before and after the i^{th} location.

negative sampling [Mikolov *et al.*, 2013b] to improve the optimization efficiency. For each location $l \in \mathbf{L}$, we sample a set of K locations that do not appear in l's context window. Then, the loss function (i.e., negative log) is defined as:

$$L_n = -\sum_{l \in \mathbf{L}} \Big(\sum_{l_c \in C(l)} (\log \sigma(\mathbf{v}_c^\top \mathbf{v}_l) + \sum_{k=1}^{K} \log \sigma(-\mathbf{v}_k^\top \mathbf{v}_l)) \Big), \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function. The K negative locations are sampled following the noise distribution $P_n(l)$, which could be the unigram distribution raised to the 3/4rd power [Mikolov *et al.*, 2013b]. It is worth noting that each location l has two latent representation vectors, one represents l as the target location, and the other represents l as the context location. Backpropagation algorithm is applied to fit the Skip-gram model.

3.2 Personalized Recommendation Model

In Section 3.1, the latent representations of locations are learned by identifying the context patterns from the global perspective, thus ignoring the personalization of individual users. To handle this issue, we propose a personalized preference learning model for personalized top-*N* location recommendations.

With the observed user-location interactions, we use $r_{u,l}$ to denote the times that a user u has visited a location l. Intuitively, as the visit count grows, we are more confident that user u likes location l. Based on such confidence, we infer users' personalized preference rankings of locations, i.e., for a given user u, the location l should be ranked higher than the location l' if $r_{u,l} > r_{u,l'}$. Note that $r_{u,l} = 0$ does not explicitly indicate u is not interested in l. It can also be caused by that u does not know l.

To devise a personalized preference learning model for top-*N* location recommendations, we adopt the WARP loss [Weston *et al.*, 2010], a pairwise ranking loss, to learn users' latent representations. By using the precision at *N* measure, the WARP loss weighs the pair-wise violations depending on the positions of locations in the ranking list. For each user u, we construct her visited location set and un-visited location set, denoted by $\mathcal{C}_u^+$ and $\mathcal{C}_u^-$ respectively. The WARP loss is defined as:

$$L_{warp} = \sum_{u \in \mathbf{U}} \sum_{l \in \mathcal{C}_u^+} L[rank(\hat{\varphi}_{u,l})], \tag{3}$$

where $rank(\hat{\varphi}_{u,l})$ is the rank of a visited location $l \in \mathcal{C}_u^+$ in u's personalized ranking list of locations. $rank(\hat{\varphi}_{u,l})$ can be estimated by $\sum_{l' \in \mathcal{C}_u^-} \mathbb{I}(\hat{\varphi}_{u,l'} \geq \hat{\varphi}_{u,l})$, where $\mathbb{I}(.)$ is the indicator function. In order to optimize the WARP loss, we replace the discrete indicator function by the continuous margin function: $max(0, 1 - \hat{\varphi}_{u,l} + \hat{\varphi}_{u,l'})$. $L(\cdot)$ transforms the rank into a loss. The implementation of $L(r)$ used in this paper is $L(r) = \sum_{i=1}^{r} \frac{1}{i}$. Note that $\hat{\varphi}_{u,l}$ indicates a user u's preference for a location l, predicted by our factorization model: $\hat{\varphi}_{u,l} = \mathbf{u}_u^\top \mathbf{v}_l$, where $\mathbf{u}_u \in \mathbb{R}^{D \times 1}$ is the latent vector of u, and $\mathbf{v}_l$ is the latent vector of l derived by the Skip-gram model in Section 3.1. In order to reconcile with the Skip-gram model, the dimensionality of users' latent representations is consistent with that of locations' latent representations.

Given huge number of locations, for most users, the un-visited locations are much more than the visited ones. In order to efficiently approximate the $rank$ function, for each user u, given a visited location l, an un-visited location l' is randomly sampled, until the one that violates the margin function. $rank(\hat{\varphi}_{u,l})$ is approximated by $\lfloor \frac{|\mathcal{C}_u^-|-1}{M} \rfloor$, where M is the number of sampling trials, $|\cdot|$ is the cardinality of a set, and $\lfloor . \rfloor$ is the floor function.

In order to better capture users' preferences for locations, we propose the C-WARP loss that extends the WARP loss to consider users' visit frequency for recommendation. Specifically, we add a weight $\theta_{l,l'}$ to each pair of positive and negative locations (l, l'). The weight is defined as: $\theta_{l,l'} = 1 + \alpha \cdot (r_{u,l} - r_{u,l'})$, where α controls the increase of the difference. Intuitively, the larger the difference, the more seriously this pair of locations violate the margin function, and hence larger weight is added to this location pair's contribution to the total loss. By considering location pair weight, the loss function is re-defined as:

$$L_{c-warp} = \sum_{u \in \mathbf{U}} \sum_{l \in \mathcal{C}_u^+} L\Big[\sum_{l' \in \mathcal{C}_u^- \cup \{\mathcal{C}_u^+ \setminus l\}} max(0, \theta_{l,l'} \cdot (1 - \hat{\varphi}_{u,l} + \hat{\varphi}_{u,l'})) \Big] + \lambda \sum_{u \in \mathbf{U}} \|\mathbf{u}_u\|^2, \tag{4}$$

where the first term is the proposed C-WARP loss, the second term is used as regularization for avoiding over-fitting, and λ controls the extent of regularization. By introducing location pair weight, during the sampling, it is not necessary to confine the negative locations to un-visited locations, the visited locations with lower visit frequency can also be sampled as negative cases.

The SGD optimization method is used to learn the latent factors of users. Specifically, we iterate through each user's visited locations and sample a negative location to update the user latent factors. The gradient of L_{c-warp} with respect to the k^{th} latent factor of u is as:

$$\frac{\partial L_{c-warp}}{\partial u_{u,k}} = L(\lfloor \frac{|\mathbf{L}|-2}{M} \rfloor)\theta_{l,l'}(v_{l',k} - v_{l,k}) + 2\lambda u_{u,k}. \quad (5)$$

The latent factor is updated as: $u_{u,k} \leftarrow u_{u,k} - \eta \frac{\partial L_{c-warp}}{\partial u_{u,k}}$, where η is the learning rate.

Once users' latent representations have been learned, we compute a user u's preference for a target location l as the inner product of the latent factor vectors $\mathbf{u}_u$ and $\mathbf{v}_l$. The location recommendations are generated by sorting the candidate locations in descending order of the predicted scores and choosing *N* top-ranked locations.

3.3 Time-aware Extension

Intuitively, users' visit behaviors strongly correlate with time. For instance, users typically visit pubs after working hours and visit shopping malls during weekends. It is thus essential to take into account temporal information to improve the location recommendation accuracy. In this section, on top of the personalized location recommendation model presented in Section 3.2, we propose a time-aware extension.

We consider two types of temporal information, i.e., hour-of-the-day and day-of-the-week. Given 24 hours per day, and 7 days per week, we have 168 time frames, denoted by $\mathbf{T}$. We assign a latent representation vector $\mathbf{w}_t \in \mathbb{R}^{D\times 1}$ to each time frame $t \in \mathbf{T}$. Note that the dimensionality of $\mathbf{w}_t$ is consistent with that of users and locations. Then, a user u's temporal preference for a location l at the time frame t is formulated as: $\hat{\varphi}_{u,l,t} = \mathbf{u}_u^\top \mathbf{v}_l + \mathbf{u}_u^\top \mathbf{w}_t + \mathbf{v}_l^\top \mathbf{w}_t$. That is, a user's preference for a location, her preference for the time of visit, as well as the correlation between the location and the time frame, are jointly modeled to produce the user's time-aware location preference. Accordingly, we modify the loss function by incorporating temporal information:

$$L^t_{c-warp} = \sum_{u\in\mathbf{U}} \sum_{l\in\mathcal{C}_u^+} L[\sum_{l'\in\{\mathcal{C}_u^+\setminus l\}} max(0, \theta_{l,l'} \cdot (1 - \hat{\varphi}_{u,l,t} + \hat{\varphi}_{u,l',t'}))] + \lambda(\sum_{u\in\mathbf{U}} \|\mathbf{u}_u\|^2 + \sum_{t\in\mathbf{T}} \|\mathbf{w}_t\|^2). \quad (6)$$

SGD is used to fit the time-aware model. The gradients of L^t_{c-warp} with respect to user u's k^{th} latent factor, time frame t's k^{th} latent factor, and time frame t''s k^{th} latent factor are computed as follows:

$$\frac{\partial L^t_{c-warp}}{\partial u_{u,k}} = L(\lfloor \frac{|\mathcal{C}_u^+|-2}{M} \rfloor)\theta_{l,l'}(v_{l',k} + w_{t',k} - v_{l,k} - w_{t,k}) + 2\lambda u_{u,k}, \quad (7)$$

Table 1: The statistics of experimental data.

	N_u	N_l	N_c
Austin	24,070	51,118	1,935,677
Chicago	13,845	37,050	486,558
Houston	11,138	29,383	512,977
Los Angeles	21,633	75,301	1,296,953

$$\frac{\partial L^t_{c-warp}}{\partial w_{t,k}} = L(\lfloor \frac{|\mathcal{C}_u^+|-2}{M} \rfloor)\theta_{l,l'}(-u_{u,k} - v_{l,k}) + 2\lambda w_{t,k}, \quad (8)$$

$$\frac{\partial L^t_{c-warp}}{\partial w_{t',k}} = L(\lfloor \frac{|\mathcal{C}_u^+|-2}{M} \rfloor)\theta_{l,l'}(u_{u,k} + v'_{l,k}) + 2\lambda w_{t',k}. \quad (9)$$

3.4 Discussion

Our personalized location recommendation model is constructed on top of several building blocks in a pipeline way. First of all, the latent representations of locations are learned through the Skip-gram model. The distributional hypothesis [Harris, 1954; Mikolov *et al.*, 2013b] shows that the locations in similar context have similar semantic meanings. This intuitively captures the relationships of users' visited locations. Secondly, our model goes beyond the global patterns learned by the Skip-gram model and realizes personalized recommendations by capturing individual users' local patterns. This is achieved by learning the latent representations of users using the proposed C-WARP loss, which refines users' preferences by capturing the visit frequency. Regarding the complexity of the proposed models, the Skip-gram model can be efficiently parallelized to cater to huge datasets, e.g., word2vec[1]. On the other hand, the complexity of optimizing the C-WARP loss can be handled by using the truncated sampling scheme [Lim and Lanckriet, 2014] in the negative location sampling process.

4 Evaluation

4.1 Experimental Settings

Data

The evaluation is conducted over real-world location-based social network data [Liu *et al.*, 2014] collected from Gowalla [2]. The data contains users' check-in information, including geographical coordinates, time stamps, etc. generated before June 1, 2011 in 4 US cities: Austin, Los Angeles, Chicago, and Houston. Table 1 summarizes the statistics of the data, where N_u, N_l, and N_c denote the number of users, locations, and check-ins respectively. Moreover, the category information of each observed location has also been collected. Locations in Gowalla are classified into 7 main categories: community, entertainment, food, nightlife, outdoors, shopping, and travel. In each main category, the locations are further classified into several subcategories.

[1]https://code.google.com/p/word2vec/

[2]http://www.yongliu.org/datasets.

Baselines

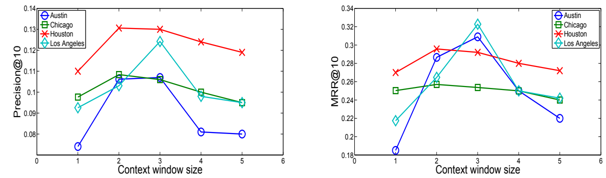
We refer to our basic model that relies on the Skip-gram model and C-WARP loss as SG-CWARP, and the time-aware extension as SG-CWARP-T. We compare our models to the state-of-the-art methods, summarized as follows: (1) *WRMF* [Hu *et al.*, 2008]. This is the weighted regularized matrix factorization model designed to handle implicit feedback data (i.e., visit a location or not) for top-N recommendation. (2) *WARP-MF* [Weston *et al.*, 2010]. This is a pairwise ranking method that utilizes matrix factorization to minimize the basic WARP loss. The latent factors of users and locations are learned by randomly sampling the positive and negative location pairs. (3) *PTMF* [Liu *et al.*, 2013b]. This approach consists of two stages. At the first stage, users' preference transitions (represented by categories of the checked-in locations) are predicted by a basic matrix factorization model, and at the second stage, users' preferences for locations in the corresponding categories (predicted by the first stage) are inferred by another matrix factorization model. Location recommendations are provided based on the category-aware preference prediction. (4) *Markov* [Gambs *et al.*, 2012]. This approach applies mobility Markov chain, where each state corresponds to a frequently visited location. A user's next movement is predicted based on her past mobility behaviors over specific temporal period and the locations she recently visited.

For time-aware location recommendations, the following 2 baselines are used: (1) *TempMF* [Gao *et al.*, 2013]. This model studies the temporal influence from the aspects of non-uniformness and consecutiveness. Matrix factorization is used to integrate the temporal influence by linearly aggregating users' preferences in different hours. (2) *Tensor*. In [Zheng *et al.*, 2010], tensor factorization is used to model the user-location-activity relations. In the experiments, we replace the activity dimension with temporal dimension and learn the latent factors for users, locations, and time frames by factorizing the tensor for time-aware location recommendations.

For each method, we use the check-in data before March 28^{th}, 2011 (around 80% of all check-ins) to train the models, and the rest data is used for testing.

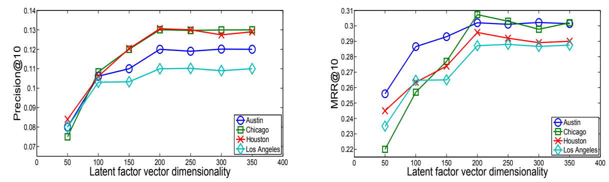
Metrics

In order to measure the quality of top-N recommendations, we use precision@N, which is the ratio of the successfully predicted locations to the top-N recommendations. Another metric we use is Mean Reciprocal Rank (MRR), a ranking metric that measures the recommendation accuracy by finding out how far from the top of the recommendation list the first successfully predicted location is: $MRR = \frac{1}{|\mathbf{U}|}\sum_{u\in\mathbf{U}}\frac{1}{R_u}$, where R_u is the position of the first successfully predicted location in the list for a user u. For time-aware recommendation, time-specific precision and MRR are calculated, and the averaged results across all time frames are reported. We use precision@N(t) and MRR(t) to represent time-aware recommendation results. Note that for latent factor models, all experiments are conducted 10 times, and the averaged results are reported. Student's t-test has been carried out to demonstrate that the results are statistically significant (two-tailed, paired t-test, p-values < 0.01).



(a) Precision. (b) MRR.

Figure 2: Performance with varying context window size (top-10 recommendations).



(a) Precision. (b) MRR.

Figure 3: Performance with varying latent factor vector dimensionality (top-10 recommendations).

4.2 Experimental Results

Design Validation

We first study how the context window size impacts the performance of our approach. The dimensionality of latent factor vectors D is set to 100 for all experiments. The context window size varies from 1 to 5. From Fig. 2, we observe the general trends are, both precision and MRR first increase with the increasing context window size; when arriving at certain threshold, the performance starts decreasing with larger context window size. Theoretically, larger window size means the target location's context can be more comprehensively modeled. However, different from large-scale text corpus where Skip-gram is typically applied, the location corpus is relatively sparse, so for our experiments, a small context window size is able to model locations' context influence.

Then, we study the influence of the dimensionality of the latent vectors. For SG-CWARP, the dimensionality is configured from 50 to 350 with 50 as the increment. From Fig. 3, we observe that more latent factors lead to higher precision and MRR. This general trend reflects that higher dimension of latent vector is capable of more accurately representing both locations and users. Nevertheless, we also observe both precision and MRR become stable after a threshold (i.e., around 200). Therefore, an optimal latent vector dimensionality can be empirically obtained such that high recommendation quality can be achieved with reasonable computational overheads. In the comparison study in the next subsection, we use 200 as the default latent vector dimensionality.

With respect to the effect of visit frequency, we compare SG-CWARP with a variant without using visit frequency (i.e., regular WARP loss is used). We observe that for different dataset, The C-WARP loss improves the performances by the

Table 2: Performance comparison on top-10 location recommendations

	Houston		Chicago		Los Angeles		Austin	
	Precision	MRR	Precision	MRR	Precision	MRR	Precision	MRR
WRMF	0.0615	0.1877	0.0621	0.1904	0.0599	0.1815	0.0566	0.1803
WARP-MF	0.0591	0.1752	0.0576	0.1750	0.0539	0.1722	0.0510	0.1764
PTMF	0.0921	0.2421	0.0966	0.2580	0.0889	0.2320	0.0810	0.2264
Markov	0.0791	0.2001	0.0805	0.2080	0.0703	0.1635	0.0668	0.1565
SG-CWARP	***0.1306***	***0.2957***	***0.1301***	***0.3073***	***0.1243***	***0.3227***	***0.1070***	***0.3090***

Table 3: Performance comparison on top-10 time-aware location recommendations

	Houston		Chicago		Los Angeles		Austin	
	Precision(t)	MRR(t)	Precision(t)	MRR(t)	Precision(t)	MRR(t)	Precision(t)	MRR(t)
TempMF	0.0688	0.1792	0.0712	0.1900	0.0665	0.1733	0.0625	0.1698
Tensor	0.0325	0.0945	0.0359	0.1022	0.0341	0.1008	0.0313	0.0886
SG-CWARP-T	***0.0940***	***0.2069***	***0.0921***	***0.2266***	***0.0870***	***0.2226***	***0.0765***	***0.2197***

percent in the range of [3.58% , 5.72%], demonstrating the effectiveness of considering visit frequency for location recommendations.

Comparison

We compare the performances of our models with that of the 6 corresponding baseline methods. We first focus on the general location recommendations without considering temporal influence. By cross-validation, for WRMF, we set the latent factor vector dimensionality, α, and regularization parameter to 150, 10, and 0.01 respectively; for PTMF, category level 2 is considered, latent factor vector dimensionality, learning rate, and regularization parameters are set to 5, 0.0001, and 0.01 respectively. For SG-CWARP, the latent factor vector dimensionality, α, regularization parameters are set to 200, 1, and 0.01 respectively; we also set optimal context window size for different datasets.

Tab. 2 summarizes the comparison results on four datasets when top-10 recommendations are provided. We observe Markov model outperforms both WARP-MF and WRMF by capturing the sequence of users' visit behaviors. However, this approach only uses the latest location for model learning but ignores the influence of other previously visited locations. PTMF also models the sequence of locations by relying on the latest location. The main difference is that semantic information (i.e., categories of locations) is considered to learn preference transition, thus improving the performance. In all cases, SG-CWARP significantly outperforms the 4 baseline models due to two main designs: (1) the Skip-gram model is applied to learn location latent representations to capture the context (before and after the target location) of users' visited locations; (2) On top of the learned location latent representations, the C-WARP loss that considers users' visit frequency is applied with a pairwise ranking algorithm to learn users' latent representations for personalized recommendations.

Next, we compare the performances of *time-aware* location recommendation models (see Tab. 3). For TempMF, we set the latent factor vector dimensionality, learning rate, user-preference parameter, location-characteristic parameter, and the time regularization parameter to 10, 0.0001, 2, 2, and 1 respectively. Although Tensor jointly learns the relationships among users, location, and time, it performs worst. This is because the original user-location matrix is already very sparse, adding another dimension of time makes the tensor even sparser, which significantly degrades the accuracy of the model. By capturing the effect of consecutiveness and non-uniformness of users' visit preferences, TempMF evidently outperforms Tensor. To summarize, by modeling pairwise interactions among users, locations, and time frames, SG-CWARP-T consistently outperforms all baselines, improving baselines by at least 29.80% and 23.14% in terms of Precision@10(t) and MRR@10(t).

5 Conclusion

In this paper, we decouple the process of jointly learning latent representations of users and locations into two separated components: learning location latent representations using the Skip-gram model, and learning user latent representations using C-WARP loss. Such a design, on the one hand, incorporates the context of locations; On the other hand, users' preferences are captured for personalized location recommendations. Furthermore, we exploit temporal information by factorizing the interactions among users, locations, and time frames. Comprehensive experiments conducted on four real world datasets demonstrate that the proposed models significantly outperform the representative methods in terms of precision and MRR. For future work, besides personalized location recommendations, we intend to apply the idea to a broader range of applications, such as user profiling, trajectory modeling, etc., to verify the effectiveness of Skip-gram model in other scenarios.

References

[Chen *et al.*, 2011] Zaiben Chen, Heng Tao Shen, and Xiaofang Zhou. Discovering popular routes from trajectories. In *Proceedings of the 27th ICDE*, 2011.

[Cheng *et al.*, 2012] Chen Cheng, Haiqin Yang, Irwin King, and Michael R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of the 26th AAAI*, 2012.

[Cheng *et al.*, 2013] Chen Cheng, Haiqin Yang, Michael R. Lyu, and Irwin King. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of 23rd IJCAI*, 2013.

[Gambs *et al.*, 2012] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, 2012.

[Gao *et al.*, 2013] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM RecSys*, 2013.

[Harris, 1954] Zellig Harris. Distributional structure. *Word*, 10(23):146–162, 1954.

[Hu and Ester, 2013] Bo Hu and Martin Ester. Spatial topic modeling in online social media for location recommendation. In *Proceedings of the 7th ACM RecSys*, 2013.

[Hu *et al.*, 2008] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 8th IEEE ICDM*, 2008.

[Kurashima *et al.*, 2013] Takeshi Kurashima, Tomoharu Iwata, Takahide Hoshide, Noriko Takaya, and Ko Fujimura. Geo topic model: Joint modeling of user's activity area and interests for location recommendation. In *Proceedings of the 6th ACM WSDM*, 2013.

[Lim and Lanckriet, 2014] Daryl Lim and Gert Lanckriet. Efficient learning of mahalanobis metrics for ranking. In *Proceedings of the 31st ICML*, pages 1980–1988, 2014.

[Liu and Xiong, 2013] Bin Liu and Hui Xiong. Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of the 2013 SDM*, 2013.

[Liu *et al.*, 2013a] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD*, 2013.

[Liu *et al.*, 2013b] Xin Liu, Yong Liu, Karl Aberer, and Chunyan Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM CIKM*, 2013.

[Liu *et al.*, 2014] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd CIKM*, 2014.

[Mathew *et al.*, 2012] Wesley Mathew, Ruben Raposo, and Bruno Martins. Predicting future locations with hidden markov models. In *Proceedings of the 2012 UbiComp*, 2012.

[Mikolov *et al.*, 2013a] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *NAACL HLT*, 2013.

[Mikolov *et al.*, 2013b] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 27th NIPS*. 2013.

[Mikolov *et al.*, 2013c] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Workshop at ICLR*, 2013.

[Sang *et al.*, 2012] Jitao Sang, Tao Mei, Jian-Tao Sun, Changsheng Xu, and Shipeng Li. Probabilistic sequential pois recommendation via check-in data. In *Proceedings of the 20th SIGSPATIAL*, 2012.

[Weston *et al.*, 2010] Jason Weston, Samy Bengio, and Nicolas Usunier. Large scale image annotation: Learning to rank with joint word-image embeddings. *Mach. Learn.*, 81(1):21–35, October 2010.

[Yang *et al.*, 2013] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhu Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, 2013.

[Ye *et al.*, 2011] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th ACM SIGIR*, 2011.

[Yuan *et al.*, 2013] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th ACM SIGIR*, 2013.

[Yuan *et al.*, 2014] Quan Yuan, Gao Cong, and Aixin Sun. Graph-based point-of-interest recommendation with geographical and temporal influences. In *Proceedings of the 23rd ACM CIKM*, 2014.

[Zhang *et al.*, 2014] Jia-Dong Zhang, Chi-Yin Chow, and Yanhua Li. Lore: Exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL*, 2014.

[Zheng *et al.*, 2010] Vincent Wenchen Zheng, Bin Cao, Yu Zheng, Xing Xie, and Qiang Yang 0001. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, 2010.