Time-aware Travel Attraction Recommendation

Kai Wang, Richong Zhang, Xudong Liu, Xiaohui Guo, Hailong Sun, and Jinpeng Huai

School of Computer Science and Engineering, Beihang University,
Beijing, 100191 China
{wangkai,zhangrc,liuxd,guoxh,sunhl,huaijp}@act.buaa.edu.cn

Abstract. The increasing number of tourists uploaded photos make it possible to discover attractive locations. Existing travel recommendation models make use of the geo-related information to infer possible locations that tourists may be interested in. However, the temporal information, such as the date and time when the photo was taken, associated with these photos are not taken into account by most of existing works. We advocate that this information give us a chance to discover the best visiting time period for each location. In this paper, we exploit a 3-way tensor to integrate context information for tourists visited locations. Based on this model, we propose a time-aware recommendation approach for travel destinations. In addition, a tensor factorization-based approach by maximizing the ranking performance measure is proposed for predicting the possible temporal-spatial correlations for tourists. The experimental results on the real tourists uploaded photos at Flickr.com show that our model outperforms existing approaches in terms of the prediction precision, ranking performance and diversity.

1 Introduction

With the boom of e-Tourism, there are emerging many online communities bringing convenience to travelers, and e-Business in tourism domain is becoming more and more flourishing. Meanwhile, tourists, especially with the widely spreading of the mobile devices, highly prefer to using such web facilities and mobile applications to enjoy their trips. In addition, travelers are likely to generate many media content, like photo, video, travelogue, etc., to share experiences in online social network. Thus, the travelers leave many online traces, which provides us expended opportunities to analyze the history travel data, so that to help more potential tourists in making decisions, such as acquainting some desired destinations, booking some suitable travel services, and planning the itinerary. However, without effective data processing methods, facing so many information, the decision-making is time-consuming and hard to reach a satisfying result. In this context, travel recommender system becomes increasingly popular and plays an important part in current tourism online services.

Existing studies on travel recommendation mainly focus on personalization, which aim to find the interaction of individual preference and features of attractions. Although the recommended items, such as destinations and tourism sites, visited by other similar tourists would make a great source of references for making the final travel decision, while the temporal information of when tourists should visit is missing. According to our statistics (presented in Section 4.1), these factors have a significant impact on the attractions choice of the tourist. For example, in late autumn, with the growing red of the maple leaves, the Fragrant Hill becomes one of the most charming places in Beijing. Then the probability that travelers would appreciate red leaves of the Fragrant Hill in late autumn should increase. Or in another case, to prove the influence of context factor, a tourist wants to know which attraction best fits his preference at night, while the recommender of ignoring the temporal context, may give an unreasonable result such as climbing Great Wall. This fact makes coherently modeling the temporal context with tourist preferences important.

The use of context-aware recommendation techniques may solve the problem of incorporating temporal context to the existing travel recommender system and building a time-aware recommendation model for predicting the most proper time period for visiting tourism sites. Several studies [1, 2] have taken into account the influence of contextual information such as the date or time for context-aware recommendations. The objective of the existing list-wise context-aware approaches is to find the best model parameters to maximize the mean average precision (MAP). However, MAP only evaluates the binary relevance of the recommended item. In practice, the preference of a user on an item is non-binary and usually evaluated by graded relevance.

To overcome the limitation of the existing context-aware recommendation approaches, in this paper, we synthesize the influence of personalization and temporal context on attraction recommendation by Tensor Factorization with a graded-degree relevance measure as the optimization objective. In particular, we exploit the nDCG, a cumulative and multilevel measure of ranking quality, as the objective function for Tensor Factorization to overcome the limitation of binary relevance measures. A gradient descent algorithm is also proposed to obtain the local optimal parameters. Moreover, we compare our proposed model with other existing matrix factorization or tensor factorization approaches on the real-life data from Flickr.com and the experimental results confirm the effectiveness of our model.

In summary, the contribution of this study is summarized as follows:

- We identify and bring to awareness the importance of temporal context for travel destination recommendation, a problem widely existing in the travel recommender systems.
- We propose the use of Tensor Factorization for temporal context-aware recommendation and utilize nDCG as the optimization objective for achieving a better ranking quality.
- We design an algorithm for predicting tourism sites that users may be interested in and the best time period for visiting. The comparative study on real-life data demonstrates the effectiveness over existing approaches.

The remainder of this paper is organized as follows. Section 2 delivers the overview of related works. Section 3 proposes the definition of our problem and the model we chose to present the interactions among travelers, attractions and temporal contexts. We also present our algorithm in this section. In Section 4, we introduce the experimental evaluation of our approach. The paper ends with brief conclusions and some discussion on future works.

2 Related Works

In order to generate appropriate recommendations and ensure the performance of recommendation systems, researchers have proposed different approaches. In this section, we present an overview of these works on travel recommendation, context-aware recommendation and learning to rank.

2.1 Travel Recommendation

With the widely development of mobile device, location-based recommendation receives more attentions from the academic circle than ever before. A large number of applications based on GPS positioning, for example, Flickr.com, a website where user can upload photos with geo-tags and timestamps, have emerged. In some studies [3–5], the timestamps of tourists uploaded photos are utilized to generate personal travel timed pathes. Then some models such as undirected graph and Markov model are adopted to recommend a travel route in given restrained conditions such as travel time and budget. Shi *et al.* [6] integrates the information of landmark categories to achieve a higher performance than basic matrix factorization and non-personalized recommendation based on popularity. However, these works do not take the context information into consideration when generating route or landmark recommendation.

Some studies claim that attractions recommendation is different from traditional collaborative filtering, such as movie or music recommendation [7, 6]. They find that almost all the travelers will visit the most popular attractions such as Forbidden City, Summer Palace when they visit Beijing. Thus, traveling histories of these attractions cannot precisely indicate their individual preferences. From another aspect, recommending less famous attractions will bring more assistance to travelers than well-known sites [6]. This characteristic of travel recommendation has also been concerned and evaluated in this study.

2.2 Context-aware Recommendation

In [8], Schmidt *et al.* define the *context* that describes as a situation or environment a device or user is in. The intuition of context-aware recommender system (CARS) entails that, in some application scenario, user preferences are not monotonous which might leads to bad performance of context-unaware recommender systems. Based on different stages of integrating the contextual information, CARS can be classify as contextual pre-filtering, contextual post-filtering and contextual modeling [2]. Recently, Tensor Factorization, as a method of contextual modeling, arouses the attention of researchers. The effectiveness of this model has been confirmed by a number of studies [9–11]. In [12], a context-aware recommender system for mobile application discovery is proposed in this study to utilize the implicit feedback of personal usage history to form a binary tensor. In our study, we advocate that binary tensor does not carry the graded degree of users' interests. So that we make use of the nDCG, a commonly used performance measure in information retrieval to characterize the user interest level on item.

2.3 Learning To Rank

As this study aims to predict the ranking of unobserved values, our task can also be viewed as a learning to rank problem. Basically, learning to rank can be classified into three types: point-wise approach, pair-wise approach and list-wise approach [13]. Most of the earlier studies focus on point-wise and pairwise approach, such as pair-wise algorithm for tag recommendation [14]. Recently, the list-wise method, which aims to minimize a list-wise loss function defined on the prediction list, usually shows a comparative outperformance to other learning to rank methods [15]. A tendency of directly optimizing the IR metric such as MAP, MRR and nDCG, has emerged in latest research on list-wise method [11, 16, 17]. The greatest challenge of learning to rank is that the non-smoothness of these metrics are not always available [18]. In this study, we present a logistic function to approximately translate the non-differentiable measures to solve the non-smoothness problem.

3 Model

3.1 Problem Definition

For the reason of taking the temporal context into account when recommending attractions to users, we extend the traditional matrix factorization model for collaborative filtering, which only considers the interaction between users and attractions, to 3-way tensor factorization model by incorporating the temporal information as another dimension besides the users and attractions.

Specifically, we denote $X \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{S}| \times |\mathcal{T}|}$ the preferences of users to attractions in some given contexts, where $\mathcal{U}$ is the set of users with $|\mathcal{U}| = m$, $\mathcal{S}$ ($|\mathcal{S}| = n$) the set of all attractions, and $\mathcal{T}$ ($|\mathcal{T}| = l$) the set of different temporal contexts. Each entry of the tensor denoted by X_{ust} indicates the degree of user $u \in \mathcal{U}$ interested in attraction $s \in \mathcal{S}$ under some temporal context $t \in \mathcal{T}$. Actually, this tensor is incomplete and noisy in practice, i.e., only a subset $O \subseteq [m] \times [n] \times [l]$ out of the whole $m \times n \times l$ entries of X is observed, and the other unobserved entries $X_{[m]\times[n]\times[l]\setminus O}$ are set to zeros.

Inspired by the family of latent factor model, such as Matrix Factorization model, we adopt CP model [19] based tensor factorization methods to learn three f-dimensional latent factor matrixes $U \in \mathbb{R}^{m \times f}$, $S \in \mathbb{R}^{n \times f}$, and $T \in \mathbb{R}^{l \times f}$ to predict the missing values. Then the entire tensor could be fitted to $\widetilde{X}$, in which each entry $\widetilde{X}_{ust}$ could be predicted through the inner product of the corresponding latent feature vectors.

$$\widetilde{X}_{ust} = \sum_{k=1}^{f} U_{uk} S_{sk} T_{tk} =< U_u, S_s, T_t > \tag{1}$$

As for the optimization objective function of these latent factors, Least Square approximations is a traditional and intuitional one, which we use as a baseline method in the experiment section, that is defined as following:

$$U^*, S^*, T^* = \underset{U,S,T}{\arg\min} \parallel X - \widetilde{X} \parallel^2. \tag{2}$$

Obviously, this loss function considers the errors for every entry of the whole tensor in an "equal odds". But for the information retrieval task, especially for the recommendation application in this work, we only highlight the precision of a few at the top of predicted results. E.g., in the Top-K recommendation which we usually show a few top-ranked items to users, the recommendation performance of the rendered items is our focus, but others with less significance.

The MAP and $nDCG$ are both rank-position sensitive and list-wise evaluation measures, but $nDCG$ reduces the contribution of the recalled items on the bottom of the returned list with a logarithmic decay factor of the position. Furthermore, different from MAP only take binary relevance, $nDCG$ can also accommodate graded, even real valued, relevance judgements.

From this point of view, we formulate our problem as to maximize the $nDCG$ values of the returned lists for all the users under all contexts. For each user $u \in \mathcal{U}$, since the value of entry $\widetilde{X}_{ust}$ represents the predicted preference of user u to attraction s under context t, the higher of this value is, the more possible the user will visit the attraction. We re-rank the attraction set according to the fitted fiber $\widetilde{X}_{u \cdot t}$ of tensor $\widetilde{X}$. Then the $nDCG$ value of the reordered list could be calculated by:

$$nDCG_{ut} = Z_{ut} \sum_{i=1}^{n} \frac{G_{uit}}{\log(1 + P_{uit})} \tag{3}$$

where P_{uit} denotes the position of i^{th} attraction in the reordered list, and $G_{uit} = 2^{R_{uit}} - 1$ represent the gain of i^{th} attraction, where $R_{uit} = X_{uit}$. Z_{ut} is the normalized factor related to $iDCG$. Then, the final $nDCG$ of tensor X can be computed as follows:

$$nDCG = \frac{1}{ml} \sum_{u=1}^{m} \sum_{t=1}^{l} nDCG_{ut} \tag{4}$$

3.2 NDCG Optimization Oriented Tensor Factorization

In this section, we will introduce how to optimize $nDCG$ based Tensor Factorization model. Firstly, a surrogate objective function approximating to $nDCG$ measure is proposed. Secondly, we present the corresponding optimization method. Finally, we present a Gradient Decent algorithm for inferring model parameters.

Based on Eq. (4), we average the $nDCG$ of all the above mentioned fibers of the entire tensor. Then, our optimization objective function can be written as

$$\mathcal{L}(U,S,T)=\frac{1}{ml}\sum_{u=1}^{m}\sum_{t=1}^{l}Z_{ut}\sum_{i=1}^{n}\frac{2^{X_{uit}}-1}{\log\left(1+(1+\sum_{j=1,j\neq i}^{n}I(\widetilde{X}_{ujt}\geq \widetilde{X}_{uit}))\right)} \quad (5)$$

where

$$I(\widetilde{X}_{ujt}\geq \widetilde{X}_{uit})=\begin{cases}1, \text{ if } \widetilde{X}_{ujt}\geq \widetilde{X}_{uit};\\ 0, \text{ else.}\end{cases} \quad (6)$$

Unfortunately, because of the non-smoothness of above indicator function, our objective function Eq. (5) is non-differentiable with respect to the variables U, S and T. Inspired by the work [20], we can smooth Eq. (6) with the logistic function as follows:

$$I(\widetilde{X}_{ujt}\geq \widetilde{X}_{uit})\approx g(\widetilde{X}_{ujt}-\widetilde{X}_{uit})=\frac{1}{1+e^{-(\widetilde{X}_{ujt}-\widetilde{X}_{uit})}}$$

Furthermore, according to [21], Maximizing Eq. (5) is equivalently minimizing following reconstructed surrogate objective function:

$$\mathcal{L}'(U,S,T)=\frac{1}{ml}\sum_{u=1}^{m}\sum_{t=1}^{l}Z_{ut}\sum_{i=1}^{n}(2^{X_{uit}}-1)\sum_{j=1,j\neq i}^{n}g(\widetilde{X}_{ujt}-\widetilde{X}_{uit}) \quad (7)$$

Ignoring the constant factor $\frac{1}{ml}$ and adding Frobenius norms of U, S, T to avoid over-fitting, the objective function to be minimized could be finally rewritten as below:

$$\begin{aligned}\mathcal{L}''(U,S,T)= &\sum_{u=1}^{m}\sum_{t=1}^{l}Z_{ut}\sum_{i=1}^{n}(2^{X_{uit}}-1)\sum_{j=1,j\neq i}^{n}g(\widetilde{X}_{ujt}-\widetilde{X}_{uit}) \\ &+\frac{\lambda}{2}(||U||^2+||S||^2+||T||^2)\end{aligned} \quad (8)$$

Then, we adopt the steepest gradient descent method to minimize the regularized surrogate objective function $\mathcal{L}''$ and derive the gradients of Eq. (8) w.r.t. the variables U, S and T respectively as following:

$$\frac{\partial_{\mathcal{L}''}}{\partial_{U_u}}=\sum_{t=1}^{l}Z_{ut}\sum_{i=1}^{n}G_{uit}\sum_{j=1,j\neq i}^{n}g'(<U_u,S_j-S_i,T_t>)[(S_j-S_i)\circ T_t]+\lambda U_u \quad (9)$$

$$\frac{\partial_{\mathcal{L}''}}{\partial_{S_s}}=\sum_{u=1}^{m}\sum_{t=1}^{l}Z_{ut}\sum_{i=1,i\neq s}^{n}(G_{uit}-G_{ust})g'(<U_u,S_s-S_i,T_t>)(U_u\circ T_t)+\lambda S_s \quad (10)$$

$$\frac{\partial_{\mathcal{L}''}}{\partial_{T_t}}=\sum_{u=1}^{m}Z_{ut}\sum_{i=1}^{n}G_{uit}\sum_{j=1,j\neq i}^{n}g'(<U_u,S_j-S_i,T_t>)[(S_j-S_i)\circ U_u]+\lambda T_t \quad (11)$$

where $\circ$ represents the element-wise product operation. With a randomized start point and suitable steps, following the negative gradient direction, objective function $\mathcal{L}''$ in Eq. (8) must gradually reach a local minimum. Algorithm 1 summarizes this steepest gradient decent based optimization method.

Algorithm 1

Input: tensor X, regularization parameter λ, learning rate η, dimension of latent factors f, max iterations $itmax$, convergence parameters $\varepsilon_u, \varepsilon_s, \varepsilon_t$
Output: Matrix U, S, T
Initialize U, S, T with random value
for $it = 1, it <= itmax, it++$ **do**
  for $u = 1, u <= m, u++$ **do**
    update $U_u = U_u - \eta \frac{\partial_{\mathcal{L}''}}{\partial_{U_u}}$ according to Eq. (9)
  end for
  for $s = 1, s <= n, s++$ **do**
    update $S_s = S_s - \eta \frac{\partial_{\mathcal{L}''}}{\partial_{S_s}}$ according to Eq. (10)
  end for
  for $t = 1, t <= l, t++$ **do**
    update $T_t = T_t - \eta \frac{\partial_{\mathcal{L}''}}{\partial_{T_t}}$ according to Eq. (11)
  end for
  if ($\| \frac{\partial_{\mathcal{L}''}}{\partial_U} \| \leq \varepsilon_u \&\& \| \frac{\partial_{\mathcal{L}''}}{\partial_S} \| \leq \varepsilon_s \&\& \| \frac{\partial_{\mathcal{L}''}}{\partial_T} \| \leq \varepsilon_t$) **then**
    break
  end if
end for
return U, S, T;

4 Experiments

In this section, we present a series of comparative experiments based on real data collected from Flickr. The results show the outperformance of the proposed method comparing with other state-of-the-art methods on recommendation.

4.1 Dataset Description

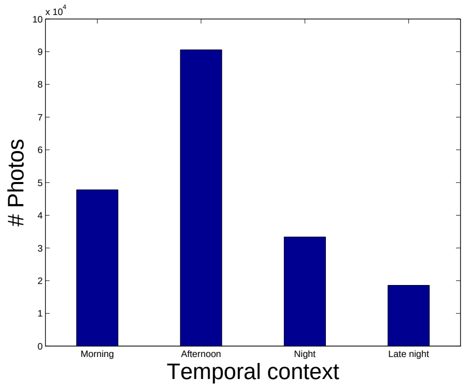
We use the public Flickr API to download 208,452 photos of Beijing, which are taken by 14,928 users. And we remove those photos without accurate taken time by judging whether the uploading time is equal or earlier than the taken time. Then, we use the geo-tags to map photos to attractions by coordinate matching and set the matching radius is set to 1000 meters. Finally, the left 180,467 effective photos are used to initialize the 3-way tensor, whose degree of sparseness is 1.822%.

We divide one day into four intervals of morning, afternoon, night and late night, and use those periods as temporal contexts. As the upload photos contain

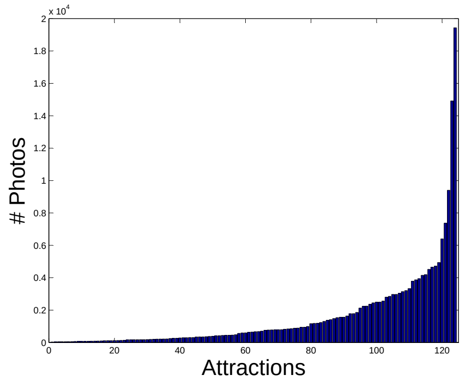
Table 1. temporal context

Context	Morning	Afternoon	Night	Late night
Period	$05:00 \sim 11:59$	$12:00 \sim 17:59$	$18:00 \sim 23:59$	$00:00 \sim 04:59$

the taken time information, we can split photos into different temporal bins. Details are shown as Table 1.



(a) photo number in each temporal context.



(b) photo number of each attractions.

Fig. 1. preliminary statistical analysis of the dataset

We conduct a preliminary analysis on photos distribution of different periods and different attractions. Fig. 1(a), obviously, reflects the following two objective and natural facts: (1)very few people travel after the midnight; (2)the photos taken in daytime by the tourists are obviously more than nighttime, and afternoon is the prime-time for visiting and taking photos. As Fig. 1(b) demonstrated, the distribution of attractions apparently takes on a long tail effect and the popularity of attractions varies widely. The top 5 popular attractions are distinctly higher than others. This phenomena conforms to the fact that hot and famous sceneries usually attract more tourists to visit. Based on this consideration, we conduct our experiments in the latter section by gradually removing those attractions to demonstrate the different characteristics between PopRec and latent factor models.

In above sections of this paper, we have claimed that the temporal distribution of visiting attractions varies dramatically. To show this fact, we investigate the proportion of photos taken at different temporal context for the most 30 famous attractive sites in Beijing. Fig.2 demonstrates that the *fraction* of uploaded photos at a specific time period over the total uploaded photos on this site is greatly influenced by the temporal contexts. This fact indicates that the different attractions show greatly diverse temporal property under the four pre-

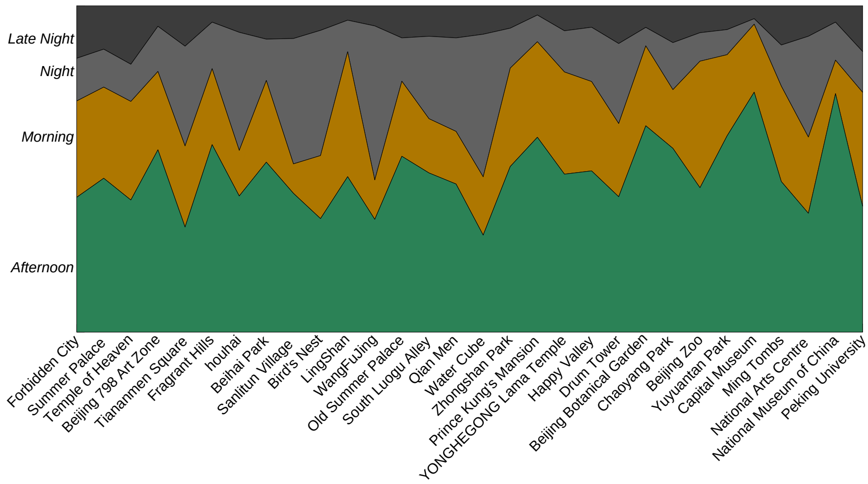


Fig. 2. The Temporal Dynamic Characteristics of Attractions

defined contexts. For instance, Capital Museum is closed at night, then the photos proportion under daytime contexts is sharply dominated. In addition, because Sanlitun Village is a famous Bar Street in Beijing and Water Cube usually is decorated with fancy neon lighting at night, both of them are with more photos at night context except afternoon. This statistical analysis supports our assumption that the temporal context is an important information for trave recommendation and it also confirms the motivation and the application potential of this work.

4.2 Baseline Methods

We conducted a series of comparative experiments with four baseline methods which are listed below to show the effectiveness of the proposed *nDCG* optimization oriented tensor factorization (*TF-nDCG*).

1. **PopRec**
   Popularity based recommendation (*PopRec*) is an intuitional and effective approach in tourism domain, since the attractions generally follow a power law distribution and people usually visit some famous sceneries. So we assume the attractions with larger number of users visited under the given temporal contexts are more popular, and recommend the top-k popular unvisited attractions. We regard this method as a non-personalized approach to compare with our latent factor based personalized model.
2. **Matrix Factorization**
   In the recommendation research works, matrix factorization (*MF*) is a basic and important benchmark model from the family of latent factor models. Because it can only handle two dimensional factors, we separate the tensor

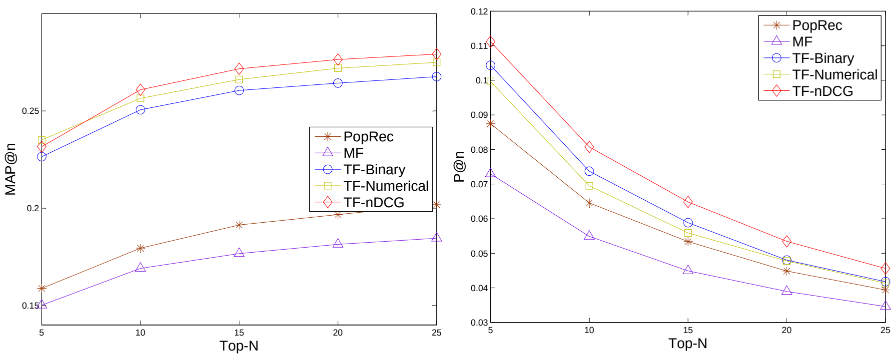
into temporal context matrix slices by fixing temporal dimension indices, then employ the regularized least squares approximation to fit each individual slice.

3. **Tensor Factorization Models**
   Regularized least square optimization for tensor decomposition as defined in Eq. (2), is a commonly used method for missing value prediction, which is essentially point-wise fitting the relevance judgements. In order to perform more subtle comparison, we feed two types of implicit feedback to these models, namely binary relevance and numerical relevance (i.e., real valued graded relevance). For the binary relevance, we set the entry X_{ust} to 0 or 1 depending on whether the user upload photos of the corresponding attraction and context or not. Correspondingly, we treat the normalized proportion of photo numbers as numerical relevance. We name these two models *TF-Binary* and *TF-Numerical* respectively.

In this study, we assume that the number of user uploaded photos indicates the degree of preference a user favors some attractive sites. This measure naturally arose to measure the user preference distribution, thus we take the proportion of user uploaded photos on a specific site over the total number of photo uploaded by this user and refer this value as the graded user preferences.

4.3 Experiment Result

Accuracy and Ranking Performance To evaluate the accuracy and ranking performance, we adopt the P@n and MAP@n as the evaluation metric respectively. We randomly choose 80% observed data of each user to form the training set, and the remaining 20% are used for evaluation. Note that we set the regularization parameter λ to be 0.01, the learning step η to be 0.01 and the number of factors d to be 5. The results are shown in Fig. 3.



(a) comparison on MAP. (b) comparison on Precision.

Fig. 3. Comparison on MAP@n and P@n

It can be observed that the performance of PopRec is much lower than other latent factor models based on Tensor Factorization that consider both personalization and context information. Overall, our TF-nDCG achieves the best performance in terms of MAP@n and P@n. Moreover, with integrating the normalized photo number, an improvement of 4% in MAP has been attained by TF-Numerical over TF-Binary. In addition, we notice that the performance of MF is even worse than PopRec, and TF-nDCG is slightly inferior (1.48%) to TF-Numerical in MAP@5.

For the characteristic of travel recommendation that travelers usually visit the most popular attractions regardless of their individual preference, the observed data are mostly concentrated on popular attractions. Past studies [7, 6] discover that travel histories in most famous attractions can not fully reflect personal preferences. Thus recommending less famous attractions is more important than recommending well-known ones. To show this less popular attraction recommendation capability, we gradually remove the most k famous attractions when generating the recommendation list of attractions.

Table 2. MAP@5 after removing k most popular attractions

k	**PopRec**	**MF**	**TF-Binary**	**TF-Numerical**	**TF-nDCG**
1	0.1345	0.1433	0.1902	0.2032	0.2141
2	0.1152	0.1305	0.1590	0.1781	0.1942
3	0.1011	0.1188	0.1390	0.1568	0.1864
4	0.0842	0.1146	0.1338	0.1459	0.1801
5	0.0784	0.1034	0.1103	0.1209	0.1698

Table 3. P@5 after removing k most popular attractions

k	**PopRec**	**MF**	**TF-Binary**	**TF-Numerical**	**TF-nDCG**
1	0.0773	0.0697	0.0908	0.0892	0.1065
2	0.0677	0.0634	0.0799	0.0811	0.0997
3	0.0632	0.0599	0.0728	0.0729	0.0948
4	0.0540	0.0579	0.0661	0.0672	0.0898
5	0.0486	0.0529	0.0588	0.0584	0.0829

The details of performance comparison after removing k popular attractions are described in Table 2 and Table 3. Due to the length limitation, we only present the comparison by MAP@5 and P@5. It can be observed that, with the increasing of number k, the performance of MF gradually becomes better than PopRec, e.g., 31.9% improvement in terms of MAP@5 and 8.85% in terms of P@5 when k is 5. Meanwhile, TF-nDCG improves the performance of MAP@5 and P@5 by 40.4% and 41.95% respectively over TF-Numeric when k is 5.

Inter-user Diversity It is not enough to merely measure the performance by accuracy metrics. Other metric, such as diversity, is also important for meeting user requirements and enhancing user experiences. To illustrate the diversification, we evaluate the inter-user diversity of each compared approaches, excluding the non-personalized method PopRec, with the metric of hamming distance. The formulation of hamming distance is defined as below.

$$H_{ijt} = 1 - \frac{S_{ij}(L_t)}{L_t} \tag{12}$$

where $S_{ij}(L_t)$ represents the number of same attractions between the recommendation list in context t of user i and user j, and L_t represents the length of the recommendation list.

We denote M the number of pairs of user, and then we can compute the overall hamming distance with Eq. (13).

$$H_{overall} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{U}\sum_{j=1,j\neq i}^{U} H_{ijt}}{T \times M} \tag{13}$$

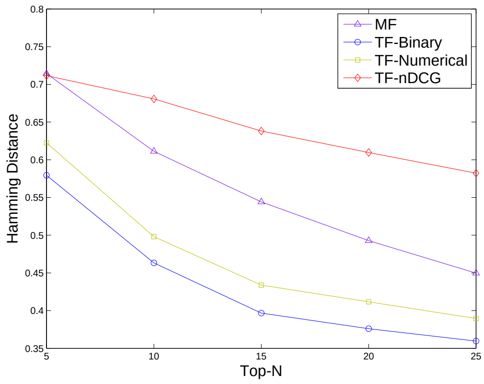


Fig. 4. comparison on diversity by hamming distance of top-n list

Fig. 4 illustrates the diversification of TF-nDCG, TF-Binary, TF-Numerical and MF. It can be seen that our model outperforms other methods in terms of the inter-user diversity. For all the other compared approaches, the diversity decreases with the increasing of N, the number of items evaluated in top-N evaluation.

In this section, we have experimentally shown the superiority of our proposed model in comparison with other commonly-used factorization techniques, i.e., matrix factorization and traditional tensor factorization. In specific, we demonstrate a concrete example of the temporal distributions of visits for attractive sites in Beijing and confirms the value of our model. In addition, we evaluate the performance of TF-nDCG in terms of accuracy, ranking performance, and diversity in a real-life dataset collected at Flickr.com.

5 Conclusion and Future Work

In this paper, we proposed a novel list-wise approach based on tensor factorization and nDCG optimization to improve the performance of time-aware recommendation for attractions. We have also presented the context-aware recommendation model and algorithm to predict the items and their best corresponding temporal context. The mathematical inference validates the effectiveness of our model. We have compared our model with other state-of-the-art algorithms and our method demonstrates a significant improvement over existing context-aware recommendation algorithms on precision, MAP and diversity.

We have not yet taken the issue of better parameter selection, such as the choosing of number of latent vectors into account. We plan to include this in our future work to increase the accuracy of travel recommendations. Moreover, our future work also includes analyzing the influence of other contextual information on attractions recommendation such as season and weather and evaluating our proposed model on other datasets. Moreover, the ranking performance metric nDCG is used as the objective function for optimization. We can investigate the feasibility of applying other metric, such as MRR, into our context-aware recommendation framework.

Acknowledgments. This work was supported partly by National Natural Science Foundation of China (No. 61103031), partly by China 863 program (No. 2012AA011203), partly by the Fundamental Research Funds for the Central Universities (No. YWF-12-RHRS-016), partly by A Foundation for the Author of National Excellent Doctoral Dissertation of PR China, partly by Beijing Nova Program and partly by Program for New Century Excellent Talents in University.

References

1. Karatzoglou, A., Amatriain, X., Baltrunas, L., Oliver, N.: Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the fourth ACM conference on Recommender systems, ACM (2010) 79–86
2. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: Recommender Systems Handbook. Springer (2011) 217–253
3. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM international conference on Information and knowledge management, ACM (2010) 579–588
4. De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C.: Automatic construction of travel itineraries using social breadcrumbs. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, ACM (2010) 35–44

5. De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C.: Constructing travel itineraries from tagged geo-temporal breadcrumbs. In: Proceedings of the 19th international conference on World wide web, ACM (2010) 1083–1084
6. Shi, Y., Serdyukov, P., Hanjalic, A., Larson, M.: Personalized landmark recommendation based on geotags from photo sharing sites. ICWSM **11** (2011) 622–625
7. Clements, M., Serdyukov, P., de Vries, A.P., Reinders, M.J.: Using flickr geotags to predict user travel behaviour. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, ACM (2010) 851–852
8. Schmidt, A., Beigl, M., Gellersen, H.W.: There is more to context than location. Computers & Graphics **23**(6) (1999) 893–901
9. Hidasi, B., Tikk, D.: Fast als-based tensor factorization for context-aware recommendation from implicit feedback. In: Machine Learning and Knowledge Discovery in Databases. Springer (2012) 67–82
10. Wermser, H., Rettinger, A., Tresp, V.: Modeling and learning context-aware recommendation scenarios using tensor decomposition. In: Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on, IEEE (2011) 137–144
11. Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Hanjalic, A., Oliver, N.: Tfmap: Optimizing map for top-n context-aware recommendation. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, ACM (2012) 155–164
12. Karatzoglou, A., Baltrunas, L., Church, K., Böhmer, M.: Climbing the app wall: enabling mobile app discovery through context-aware recommendations. In: Proceedings of the 21st ACM international conference on Information and knowledge management. CIKM '12, New York, NY, USA, ACM (2012) 2527–2530
13. Cao, Z., Qin, T., Liu, T.Y., Tsai, M.F., Li, H.: Learning to rank: from pairwise approach to listwise approach. In: Proceedings of the 24th international conference on Machine learning, ACM (2007) 129–136
14. Rendle, S., Schmidt-Thieme, L.: Pairwise interaction tensor factorization for personalized tag recommendation. In: Proceedings of the third ACM international conference on Web search and data mining, ACM (2010) 81–90
15. Xia, F., Liu, T.Y., Wang, J., Zhang, W., Li, H.: Listwise approach to learning to rank: theory and algorithm. In: Proceedings of the 25th international conference on Machine learning, ACM (2008) 1192–1199
16. Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Oliver, N., Hanjalic, A.: Climf: learning to maximize reciprocal rank with collaborative less-is-more filtering. In: Proceedings of the sixth ACM conference on Recommender systems, ACM (2012) 139–146
17. Taylor, M., Guiver, J., Robertson, S., Minka, T.: Softrank: optimizing non-smooth rank metrics. In: Proceedings of the international conference on Web search and web data mining, ACM (2008) 77–86
18. Quoc, C., Le, V.: Learning to rank with nonsmooth cost functions. Proceedings of the Advances in Neural Information Processing Systems **19** (2007) 193–200
19. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM review **51**(3) (2009) 455–500
20. Chapelle, O., Wu, M.: Gradient descent optimization of smoothed information retrieval metrics. Information retrieval **13**(3) (2010) 216–235
21. Valizadegan, H., Jin, R., Zhang, R., Mao, J.: Learning to rank by optimizing ndcg measure. Advances in Neural Information Processing Systems **22** (2009) 1883–1891