Exploiting Visual Content for Travel Location Recommendation

Thaair Ameen
dept. Presidency of Mosul University
Mosul University,
Mosul, Iraq
thaearm@gmail.com

Mustafa Khalid
dept. College of Control Science and Engineering
Zhejiang University
Hangzhou, China
mustafa_khalid@zju.edu.cn

Asmaa Waleed Abdulqahar
dept. College of Medicine
Al-Mustansiriya University
Baghdad, Iraq
Asmaaalkubaissy2@gmail.com

Ali Tariq
dept. College Of Dentistry
Mosul University
Mosul,Iraq
ali.khalil987@uomosul.edu.iq

2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) | 978-1-6654-4231-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICECET52533.2021.9698444

***Abstract*—Image has become an important source to improve the quality of travel location recommendation, it reflects user interests and travel location properties. However, little work exists for travel location recommendation by exploiting images. In this study, we propose a method for accurate and personalized travel location recommendations using visual content. Specifically, a convolutional neural network is used to extract visual content which is used to learn the latent feature representation of implicit feedback and uncover user interests and travel location properties. In addition, the third-party web service is used to extract age and gender features from the images to understanding user interest and the travel location properties. Experimental results on real-world datasets demonstrate the effectiveness of our method.**

Keywords—Travel location recommendation, matrix factorization, and visual content, classification.

I. INTRODUCTION

With the rapid development of electronic devices and ubiquitous Internet access, community-contributed geo-tagged images (CCGI) have become common on social media sites (such as Flickr). These CCGIs contain a wealth of information, which can remind users of their interest in travel locations and provide recommended tasks.

Recently, two research lines have been carried out to expand the above-mentioned past usage data, which will be briefly reviewed below.

In the first-line research, many methods use a third-party web service (T-PWS) to classify gender and age features, which helps describe user interests [2, 3] and travel location properties [4]. Certainly, there are some locations that are preferred by males and females where they would like to travel. For example, females may prefer to visit shopping arcades while men would like to pay a visit to the boxing ring. Further, the age of the user is also useful for determining their travel location, because certain travel locations are more likely to be chosen by a specific age group e.g., teenagers may prefer to go to animation and manga museums. Some public websites (i.e., www.alchemyapi.com) can mine information from the visual content of the image and provide gender and age features of all faces in the image. By collecting statistics of the features extracted from these images acquired from the travel location, we can obtain the gender and age distribution of that particular travel location. Similarly, we can generate statistics on the features extracted from the images taken by the users. (T-PWS) has reasonable performance (the accuracy rate in determining the age and gender of the face in the Adience benchmark [5] image is more than 88%, and the state-of-the-art work [6, 7, 8, 9, and 10].

In the second-line research, most recommendation methods only consider the impact of social relevance from users who use heterogeneous CCGI data sets [11, 12]. Therefore, with the development of computer vision, it is worthwhile to study visual features from images to reveal the potential of visual content [13], because images reflect rich features about user interests and travel location properties.

In this study, extending from the above two research directions, we aim to introduce a new method to use visual content for travel location recommendation (VCR). Convolutional Neural Network (CNN) is used to extract visual content from geotagged images and incorporate it into a weighted matrix factorization (WMF) method to learn the latent factor representation of users and travel locations to capture two types of correlations user-user and travel location-travel location social relevance. In addition, we use (T-PWS) to extract more enhanced features (i.e., gender and age) to make the statistics for more additional features. The main contributions of this research are summarized as follows.

- Propose a hybrid recommendation method for travel location recommendations by fusing CNN with T-PWS extractions features to devise a method for a given user to visit a travel location.
- Show that various relations can be extracted from linked CCGIs data based on visual content to provide a way to enhance user interest and travel location properties.
- We conduct experiments on CCGIs dataset to evaluate the effectiveness of the proposed method

II. PRELIMINARIES

Definition 1 (Geotagged image): A geotagged image indicates of the location embedded into images. Can be defined as a tuple $p = (id, u, g, \mathrm{X}, t)$, which contain a image's

978-1-6654-4231-2/21/$31.00 ©2021 IEEE

unique identification id and coordinates X is taken by user u at time t.

Definition 2 (User visit): interest from user u to travel location l at the time t is represent as a tuple $\beta = (u, l, t)$.

Definition 3 (Travel location): A travel location l is a specific interest location that visited by a user.

Definition 4 (Image collection): A collection of images is to obtain information to keep on record is represented as a set $\mathcal{P} = \{\mathcal{P}_{u_1}, \mathcal{P}_{u_2}, \dots, \mathcal{P}_{u_n}\}$, where $\mathcal{P}_{u_i}$ is denote to the collection of images by user u_i.

Our research problem can be formulated as: Given N users, M travel locations, images of CCGIs P, we aim to recommend top-N travel location to each user.

III. METHODOLOGY

The framework of our method is illustrated in Fig. 1 There are three objects, in our studied problem, namely, users, travel locations, and images. Let $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ be the set of N users, $\mathcal{L} = \{l_1, l_2, \dots, l_M\}$ be the set of M travel locations. Then, from the visiting records; we build the original user-travel location matrix $\boldsymbol{R} \in \mathbb{R}^{N \times M}$. We use $\mathbf{u}_i$ denotes the user latent factor vector of u_i, and $\mathbf{l}_j$ denotes the travel location latent factor vector of l_j. Images set $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_D\}$, where D include total number of images, which reflect users' interest and travel locations properties. However, CNN use to process the content of images to train $\mathbf{u}_i$ and $\mathbf{l}_j$ for social correlations; and T-PWS are mined to exploit auxiliary features (i.e., age, and gender), which is use to understanding a user interest and travel location properties, we introduce a WMF as a state-of-the-art method that has been proven to be efficient and effective in travel location recommendation [14]. Thus, in this study, visual content incorporated into WMF for travel location recommendation.

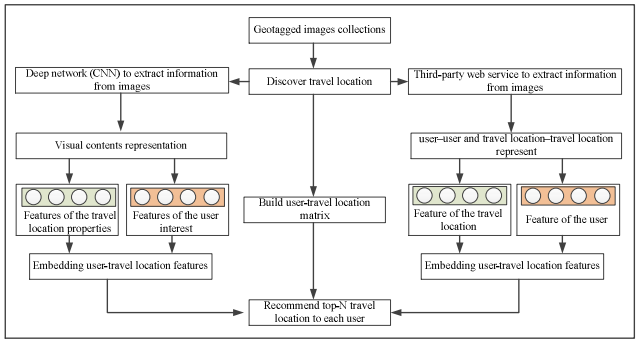

Fig. 1. Framework of the proposed method: CNN part in the left side; third-party Web service part in the right side.

A. DISCOVERING TRAVEL LOCATIONS

In order to find the travel location of highly users in his visit records, we use P-DBSCAN [15] to cluster images using their spatial proximity. The input is a set of images $\mathcal{P}$, and output of a P-DBSCAN is a set of travel locations $\mathcal{L} = \{l_1, l_2, \dots, l_M\}$. Each element $l = (\mathcal{P}_l, g_l)$, where $\mathcal{P}_l$ is a cluster of images and g_l is the geographical coordinates of the centroid of the cluster g_l and is computed from group of geotags annotated to images in the cluster $\mathcal{P}_l$.

B. IMPLICIT INFORMATION MODELING

The CCGI metadata contains play counts per user and per travel location, which is from past usage data, to build user-travel location matrix, Therefore, these matrix r_{ij} implicit construct (u, l, t), which means that user u visits travel location l at time t. To calculate the interest from user u_i to l_j, we insert the weighted effect into WMF algorithm [14]. For each user-travel location pair, we introduce the interest $\mathbf{P}_{ij}$ values which is obtained from binarizing the r_{ij}, as shown in Equation (1), and $\mathbf{C}_{ij}$ value is reflect our confidence in observing $\mathbf{P}_{ij}$, and α, ϵ are hyper parameters as shown in Equation (2).

$$\mathbf{P}_{ij} = \begin{cases} 1 \; r_{ij} > 0 \\ 0 \; r_{ij} = 0 \end{cases} \tag{1}$$

$$\mathbf{C}_{ij} = 1 + \alpha \, log\left(1 + \frac{r_{ij}}{\epsilon}\right), \tag{2}$$

C. EXTRACTING VISUAL CONTENTS USING THIRD-PARTY WEB SERVICE

To build a recommendation system by extracting the facial features based on age and gender classification , which can subsequently indicate the users preference and the properties of the travel location in order to predict if the user is interested in that particular travel location or not (e.g., as compared to males, females may prefer to visit the shopping places). When a new user enters the system, our classifier will provide a binary value of the travel locations that the user likes or dislikes. Accordingly, we can recommend the travel location to the user. User's features, such as, age and gender represent user interests and travel location properties (e.g., predicting a male's face in a travel location means that males are more interested in that particular travel location). Based on this input, our classifier will generate a binary value to represent the user's likes or dislikes. Thereafter we can recommend travel locations to users. This information is used to build a profile of each user and travel location l_j, i.e., $E_{u_i} = (f_1^{u_i}, f_2^{u_i}, \dots, f_F^{u_i})$ and $E_{l_j} = (f_1^{l_j}, f_2^{l_j}, \dots, f_G^{l_j})$, where F and G are the number of features for both $\mathcal{U}$ and $\mathcal{L}$, respectively. The details of explicit information from mining CCGIs dataset are shown in Table 1.

Based on this input, our classifier will generate a binary value to represent the user's likes and dislikes. Thereafter, we can recommend travel location to the user. This explicit information is used to build a profile of each user and travel location l_j, i.e., $E_{u_i} = (f_1^{u_i}, f_2^{u_i}, \dots, f_F^{u_i})$ and $E_{l_j} = (f_1^{l_j}, f_2^{l_j}, \dots, f_G^{l_j})$, where F and G are the number of features for both $\mathcal{U}$ and $\mathcal{L}$, respectively. The details of explicit information from mining CCGIs dataset are shown in Table 1.

TABLE I. THE LIST OF THE EXPLICIT INFORMATION OF USER AND TRAVEL LOCATION.

The information of User and travel Location	Description
UserMaleFaceNum	The total number of male faces in all the user's images.
UserFemaleFaceNum	The total number of female faces in all the user's images.

Identify applicable funding agency here. If none, delete this text box.

UserYouthNum UserTeenagerNum UserElderNum UserMiddleAgeNum	The number of youth, teenager, elder and middle-age faces in all the user's images.
UserSeqRatio	The ratio between the total number of sequences made by the user and the maximum number of sequences for a user.
LocUserNum	The number of users who visited the travel location.
LocCatNum_Shopping LocCatNum_Education LocCatNum_Religious LocCatNum_Food LocCatNum_Transportation LocCatNum_Entertainment LocCatNum_Cultural	The number of travel locations associated with Shopping, Education, Religious, Food, Transportation, Entertainment, and Cultural.
LocVisitedRatio	The ratio between the number of users currently visiting the travel location and the total number of users
LocimgRatio	The ratio between the number of images for the travel location and the maximum number of images for a travel location.
LocMaleFaceNum	The number of male faces in all images that are taken at the location.
LocFemaleFaceNum	The number of female faces in all images that are taken at the location.
LocYouthNum LocTeenagerNum LocElderNum LocMiddle-AgeNum	The number of youth, teenager, elder and middle-age faces in all images that are taken at the travel location.

By popularizing the achievements in the field of artificial intelligence based on deep learning, T-PWS will promote a new generation of keen applications for understanding computer vision. It is an easy to use, high performance T-PWS for real time computer vision that can provide enterprises with the intelligence needed to transform large amounts of unstructured data into business driven activities. However, the effect of T-PWS based on face extraction of user interest may have different advantages. Let S_{ik} be the strength of relations between two users (e.g., u_i and u_k), such that S_{ik} indicates the similarity of two user interests. Therefore, we use the cosine rating similarity to calculate S_{ik} as follows:

$$S_{ik} = \frac{\sum_{g=1}^{z} f_{ig} \cdot f_{kg}}{\sqrt{\sum_{g=1}^{z} f_{ig}^2} \cdot \sqrt{\sum_{g=1}^{z} f_{kg}^2}}, \tag{3}$$

where f_{ig} and f_{kg} represent the g-th feature of users u_i and u_k, respectively; and z is the number of features. We also use S_{jk} to indicate the similarity of two travel location properties. Then, the strength of connections between two travel locations (e.g., l_j and l_k) can be calculated as follows:

$$S_{jk} = \frac{\sum_{g=1}^{\hat{z}} f_{jg} \cdot f_{kg}}{\sqrt{\sum_{g=1}^{z} f_{jg}^2} \cdot \sqrt{\sum_{g=1}^{z} f_{kg}^2}}, \tag{4}$$

where f_{jg} and f_{kg} represent the g-th feature of travel locations l_j and l_k, respectively; and $\hat{z}$ is the number of features

D. EXTRACTING VISUAL CONTENTS USING CONVOLUTIONAL NEURAL NETWORKS

The For computing the potential of images, we employ the state-of-the art CNN architecture VGG-16 [16] to extract meaningful features from the images, which consists of 16 layers, 13 convolutions, 3 fully connected, 5 max pooling and one softmax layer. The size of images is resized to $224 \times 224 \times 3$ as input to CNN, where 3 is the number of channels i.e., RGB.We use pretraing to initialize the whight of VGG-16 on place database[1]that contains 7,076,580 images from 476 scene categories. The output of each image is termed as visual content which is a vector of dimension $d = 4096$.

E. MODELING USER-INTEREST AND TRAVEL LOCATION-PROPERTY

The content that users are interested in is also embedded in the images on CCGI. The image contains visual content that reflects the user's interest in the travel location, that is, social relevance. It can help solve the sparsity problem of user-travel location interaction to a certain extent because insufficient observation of user visit behavior can be compensated by inferring user interest through observed image behavior. Therefore, we recommend using the features in the image to improve the learning of the user's latent interests, as shown below.

$$\min \frac{1}{2} \sum_{i=1}^{N} \sum_{p_k \in \mathcal{P}_{u_i}} (\mathbf{F}_{ik} - \mathbf{u}_i CNN(p_k))^2, \tag{5}$$

where $p_k \in \mathbb{R}^{K \times d}$ represents the image latent factor vector in user-interest content, and d is the dimension of the visual contents.

Similarly, images from the content of the travel location property can also be used to learn the travel location properties, as shown below.

$$\min \frac{1}{2} \sum_{j}^{M} \sum_{p_k \in \mathcal{P}_{l_j}} \left(\mathbf{H}_{jk} - \mathbf{l}_j CNN(p_k')\right)^2, \tag{6}$$

where $p_k' \in \mathbb{R}^{K \times d}$ represents the image latent factor vector in location-property content. Both p_k and p_k' represent image visual contents, where the former is in user visual context related to user-interest content, and the latter is in travel location context related to travel location-property content. Thus, we expect these two visual contents to be different but with certain overlaps, and propose a ℓ-1 norm to capture such relationship,

$$\min \|p_k - p_k'\|_1 \tag{7}$$

F. TRAVEL LOCATION RECOMMENDATION FRAMEWORK

In this work, we introduce our solutions to mathematically determine CNN and T-PWS extractions into hybrid method. Thus, we propose a VCR travel location recommendation method to exploit the visual content and age,

[1]The places database and pretrained networks are available at http://places.csail.mit.edu

gender features simultaneously. The proposed unified method is used to solve the following optimization problem:

$$\begin{aligned}\mathcal{J} = \min_{\mathbf{U},\mathbf{L},\mathbf{F},\mathbf{H}} & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{C}_{ij}\left(\mathbf{P}_{ij}-\mathbf{u}_i^T\mathbf{l}_j\right)^2 \\ & +\frac{\lambda_1}{2}\left(\sum_{i=1}^{n}\sum_{k\in K(i)}\mathbf{S}_{ik}\left\|\mathbf{u}_i-\mathbf{u}_k\right\|_F^2\right. \\ & \left.+\sum_{j=1}^{m}\sum_{k\in K^{\wedge}(j)}\mathbf{S}_{jk}\left\|\mathbf{l}_j-\mathbf{l}_k\right\|_F^2\right) \\ & +\frac{\lambda_3}{2}\left(\sum_{i=1}^{n}\sum_{p_s\in\mathcal{P}_{u_i}}\left\|\mathbf{A}_{is}-\mathbf{u}_i^T\cdot\mathbf{F}\cdot CNN(p_s)\right\|_F^2\right. \\ & \left.+\sum_{j=1}^{m}\sum_{p_t\in\mathcal{P}_{l_j}}\left\|\mathbf{B}_{jt}-\mathbf{l}_j^T\cdot\mathbf{H}\cdot CNN(p_t)\right\|_F^2\right) \\ & +\frac{\lambda_2}{2}\left(\|\mathbf{U}\|_F^2+\|\mathbf{L}\|_F^2\right) \\ & +\frac{\lambda_4}{2}\left(\|\mathbf{F}\|_F^2+\|\mathbf{H}\|_F^2\right),\end{aligned} \tag{8}$$

Can reach a harmonic status when $\mathcal{J}$ converges in the learning phase. The key idea is to minimize the objective function $\mathcal{J}$ by fixing one variable the other variables. We use gradient descent updating rules to learn the latent parameters and correlation matrixes (see Eq. 9-13).

$$\begin{aligned}\mathbf{u}_i \leftarrow \mathbf{u}_i+\alpha\left(\Delta_{ij}\mathbf{l}_j-\lambda_1\sum_{g\in G(i)}\mathbf{S}_{ik}(\mathbf{u}_i-\mathbf{u}_k)\right. \\ +\lambda_3\sum_{p_s\in\mathcal{P}_{u_i}}\left(\mathbf{A}_{is}-\mathbf{u}_i^T.\mathbf{F}.CNN(p_s)\right)\mathbf{F}.CNN(p_s)^{\mathrm{T}} \\ \left.-\lambda_2\mathbf{u}_i\right)\end{aligned} \tag{9}$$

$$\begin{aligned}\mathbf{l}_j \leftarrow \mathbf{l}_j+\alpha\left(\Delta_{ij}\mathbf{u}_i-\lambda_1\sum_{q\in Q(j)}\mathbf{S}_{jk}\left(\mathbf{l}_j-\mathbf{l}_k\right)\right. \\ +\lambda_3\sum_{p_t\in\mathcal{P}_{l_j}}\left(\mathbf{B}_{jt}-\mathbf{l}_j^T.\mathbf{H}.CNN(p_t)\right)\mathbf{H}.CNN(p_t)^{\mathrm{T}} \\ \left.-\lambda_2\mathbf{l}_j\right)\end{aligned} \tag{10}$$

$$\begin{aligned}\mathbf{F} \leftarrow \mathbf{F}+\alpha\left(\lambda_3\sum_{p_t\in\mathcal{P}_{u_i}}\left(\mathbf{A}_{is}-\mathbf{u}_i^T\cdot\mathbf{F}\cdot CNN(p_s)\right)\mathbf{u}_i\cdot CNN(p_s)^{\mathrm{T}}\right. \\ \left.-\lambda_4\mathbf{F}\right)\end{aligned} \tag{11}$$

$$\begin{aligned}\mathbf{H} \leftarrow \mathbf{H}+\alpha\left(\lambda_3\sum_{p_t\in\mathcal{P}_{l_j}}\left(\mathbf{B}_{jt}-\mathbf{l}_j^T\cdot\mathbf{H}\cdot CNN(p_t)\right)\mathbf{l}_j\cdot CNN(p_t)^{\mathrm{T}}\right. \\ \left.-\lambda_4\mathbf{H}\right)\end{aligned} \tag{12}$$

$$\Delta_{ij}=\mathbf{P}_{ij}-\mathbf{u}_i^T\mathbf{l}_j \tag{13}$$

With the abovementioned update rules, the algorithm of VCCI is summarized in Algorithm 1. In line 1, we initialize the weight of VGG-16 on the place database. In line 2, we randomly initialize **U**, **L**, **F** and **H**. From line 3 to line 8, we update the parameters until convergence. Finally, for each user u_i, we sort the score $\mathbf{u}_i^T\mathbf{l}_j$, and recommend the travel locations with the highest scores.

ALGORITHIM 1 The proposed travel locations recommendation frame work

Input: The rating information **P**, the social information **T**, the number of latent factors k and α
Output: The user preferences matrix **U** and the travel location characteristic matric **L**
1: Initialize **U**, **L**, **H** and **F** randomly
2: Construct S From P and T
3: **while** Not convergent do
4: Calculate $\frac{\partial\mathcal{J}}{\partial\mathrm{U}}$, $\frac{\partial\mathcal{J}}{\partial\mathrm{L}}$, $\frac{\partial\mathcal{J}}{\partial\mathrm{F}}$ and $\frac{\partial\mathcal{J}}{\partial\mathrm{H}}$
5: Update $\mathbf{U}\leftarrow\mathbf{U}-\gamma_u\frac{\partial\mathcal{J}}{\partial\mathbf{U}}$
6: Update $\mathbf{L}\leftarrow\mathbf{L}-\gamma_l\frac{\partial\mathcal{J}}{\partial\mathbf{L}}$
7: Update $\mathbf{H}\leftarrow\mathbf{H}-\gamma_h\frac{\partial\mathcal{J}}{\partial\mathbf{H}}$
8: Update $\mathbf{F}\leftarrow\mathbf{F}-\gamma_m\frac{\partial\mathcal{J}}{\partial\mathbf{M}}$
9: **end while**
10: return $\mathbf{P}=\mathbf{u}_i^T\mathbf{l}_j$

IV. EXPERIMENTS

In this section, we evaluate the following: (1) how is the proposed method comparison with other state-of-the-art recommendation methods? (2) how does the convolutional neural network and third-party web service Impacts of the proposed method.

A. DATASET

We conduct experiments on the public CCGIs dataset [17], which comes from 7,387 users. This dataset is composed of heterogeneous past usage data related to image albums [18], and nine popular tourist cities (i.e., New York, Los Angeles, Chicago, Barcelona, Berlin, London, Paris, Rome, and San Francisco) are used to evaluate us Recommend system performance. We deleted "selfie images" and images without latitude and longitude to reduce information interference. The final statistics of the data set are shown in Table 2.

TABLE II. STATISTICS OF THE CCGI DATASET.

Cities	**Users**	**Travel locations**	**Images**		**Ratings**
			(Filtere)	**(Row)**	
Barcelona	217	23	5,853	15,704	218
Berlin	209	22	11,083	13,420	209
Chicago	321	35	12,304	22,104	321
London	923	46	14,256	43,557	923
Los Angeles	134	20	3,961	10,122	34
New York	620	44	12,049	34,374	610
Paris	746	36	10,879	24,507	746
Rome	323	24	7,828	18,416	323
SanFrancisco	466	34	10,060	24,572	466
Total	3,959	284	88,273	206,776	3,850

B. PARAMETER SETTINGS

In this section, we mainly give the meaning and settings of several parameters used in our experiments.

- Parameter employed for P-DBSCAN with adaptive density: To discover travel locations from CCGIs, based on the work of Kisilevich et al[15], the minimum number of users Min Owners, and density drop threshold w were empirically set to 50, 50, and 0., respectively.

- Visiting stay threshold $visit_{thr}$. To obtain the duration time of a visitor when he/she travelled a location, $visit_{thr}$ was set 6h.
- Regularization term parameter $\lambda_1, \lambda_2,$: this parameter controlled the contribution of each part during collaborative decomposition, for matrix factorization based methods in the following experiments λ_1, λ_2 were empirically set to 0.001.

In our experiment, we randomly select x% of all travel locations visited by each user as the training set, and the remaining 1-x% as the test set, where x is diverse {60, 80}. We also deleted images related to the tagging in the test data to ensure that the test data information will not be exposed during the training process. Research on the ability of the proposed framework to deal with data sparse problems uses two widely used evaluation indicators, namely precision@N and recall@N. N is set to 5 and 10. Our evaluation indicators are as follows:

$$precision@N = \frac{\sum_{u_i \in \mathcal{U}} |TopN(u_i) \cap M(u_i)|}{\sum_{u_i \in \mathcal{U}} |TopN(u_i)|}, \quad (14)$$

$$recall@N = \frac{\sum_{u_i \in \mathcal{U}} |TopN(u_i) \cap M(u_i)|}{\sum_{u_i \in \mathcal{U}} |M(u_i)|}, \quad (15)$$

commended for user u_i in the training set. $M(u_i)$, represents a set of visited travel locations in the testing set.

C. PERFORMANCE COMPARISON

To answer the first research question, we compare our method with the following representative methods.

- User-based collaborative filtering **(UCF):** [19] introduced UCF, which was a classical memory-based recommendation method that uses the cosine similarity between users on the user–travel location matrix to generate the prediction scores.
- Location recommendation with temporal effect **(LRT):** [20] introduced LRT, which is a time-enhanced matrix factorization method, in which a user's interest in travel locations drifts over time. LRT models each user by using different latent features for various time slots and then computes all the latent features for the travel location recommendation task.
- Geographical probabilistic factor model **GeoPFM:** [21] introduced GeoPFM, which is based on that user's interest preference and geographical influence are alternately affected. Thus, GeoPFM jointly learns the interest preference and geographical influence of users.
- Weighted matrix factorization **(WMF):** [22] introduced WMF, which considers visited and unvisited travel locations as positive and negative examples, respectively, and assigns more weight to the positive examples than the negative ones. WMF is our basic heterogeneous travel location recommendation method, as defined in Eq. (5), without considering the images.
- Visual content enhanced POI recommendation (VPOI): [23] introduced VPOI, which captures the heterogeneous latent feature and extracts the visual content from images to train the heterogeneous latent feature representation for the personalized travel location recommendation task. The difference with our method is that VPOI only uses images to learn the latent feature vector representations.

The results are listed in Table 3, and the following observations are presented:

- The performances of all methods steadily increase with the increase of training set.
- VCR performs better compared to other baseline methods. That is because VCR combines implicit information, T-PWS extraction, and CNN extraction, which means VCR is more robust to the data sparsity problem.
- The proposed framework significantly outperforms VPOI. That is because of the incorporating of CNN extraction (i.e., visual content) and T-PWS extraction (i.e., age and gender), while VPOI ignores explicit information to mitigate the sparsity problem.

With the abovementioned analysis, we can draw an answer to the first questions with the exploit of deep network extraction and T-PWS extraction, VCR not only improve the travel location recommendation performance compared with other state-of-the-art but also can mitigate the data sparsity problem in travel location recommendation methods.

D. IMPACTS OF CONVOLUTIONAL NEURAL NETWORK EXTRACTION AND THIRD-PARTY WEB SERVICE EXTRACTION

In this section, we investigate the impacts of T-PWS and CNN extractions on VCR. In detail, we eliminate the impact of T-PWS and CNN by defining the following two variants of VCR.

TABLE III: RECOMMENDATION COMPARISONS IN TERMS OF PRECISION AND RECALL

x%	Metric	(a) UCF	(b) LRT	(c) GeoPFM	(d) WMF	(e) VPOI	(f) VCR	Improvement f vs. best	Improvement f vs. d
(60%)	precision@5	0.0351	0.0487	0.0737	0.0752	0.0831	**0.0911**	8.78%	17.45%
	recall@5	0.0300	0.0318	0.0418	0.0596	0.0690	**0.0848**	18.63%	29.72%
(80%)	precision@5	0.0428	0.0513	0.0786	0.0862	0.0901	**0.0958**	5.95%	10.02%
	recall@5	0.0328	0.0277	0.0537	0.0642	0.0765	**0.0885**	13.56%	27.46%
(60%)	precision@10	0.0269	0.0312	0.0369	0.0524	0.0619	**0.0735**	15.78%	28.71%
	recall@10	0.0293	0.0347	0.0429	0.0620	0.0650	**0.0790**	17.72%	21.52%
(80%)	precision@10	0.0277	0.0467	0.0387	0.0573	0.0658	**0.0816**	19.36%	29.78%
	recall@10	0.0374	0.0693	0.0576	0.0658	0.0681	**0.0850**	19.88%	22.59%

- VCR\T-PWS - Eliminating the impact of T-PWS (i.e., age and gender) classification by setting λ_1=0 in Eq. (8).

- VCR\CNN - Eliminating the impact of visual content by setting λ_3=0 in Eq. (8).
- VCR\T-PWS\CNN - Eliminating the impacts of both visual content and context information by setting λ_1, λ_2, λ_3 and λ_4=0 in Eq. (8).

The comparison results are shown in Table 4 for precision@N and recall@N, respectively. Note that we only show the results with 80% and √ is use to indicate of matching features and × otherwise. When eliminating the impact of T-PWS (i.e., age and gender) from the proposed framework, the performance of VCR\CNN degrades. We have the similar observation for VCR\T-PWS when eliminating the impact of visual content. Eliminating the impacts of both CNN and T-PWS, VCR\CNN\T-PWS obtains worse performance than both VCR\CNN and VCR\T-PWS, suggesting that visual content and T-PWS contain complementary information to each other for recommendation.

V. CONCLUSIONS AND FUTURE WORK

TABLE IV. THE IMPACTS OF CONVOLUTIONAL NEURAL NETWORK AND THIRD-PARTY WEB SERVICE EXTRACTIONS.

Features	T-PWS	CNN	precision @5	recall @5	precision @10	recall @10
None	×	×	0.0126	0.0177	0.0111	0.0310
T-PWS	✓	×	0.0130	0.0182	0.0115	0.0323
CNN	×	✓	0.0164	0.0230	0.0145	0.0416
T-PWS+CNN	✓	✓	0.0172	0.0242	0.0154	0.0433

In this study, we investigate how to exploit visual contents on CCGIs. To utilize CCGIs effectively, we use a CNN to extract visual content from images and use such content to learn the latent feature representation of implicit feedback and uncover social correlations. We also use the T-PWS to extract context features (i.e., age and gender) and understand T-PWS affect, thereby leading to a unified travel location recommendation framework VCR. Experimental results on a real-world dataset demonstrate the effectiveness of our method. This work can be extended in the future with the following directions: (i) the latent feature representation for a new travel location can be estimated from its images to mitigate the travel location cold-start problem. (ii) Additional information (e.g., textual content medications) can be incorporated to improve recommendation performance.

REFERENCES

[1] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation, In SIGKDD. 2014, 831-840.

[2] Q. You, S. Bhatia, T. Sun, and J. Luo, The eyes of the beholder: Gender prediction using images posted in online social networks, In IEEE International Conference on Data Mining Workshop. 2014, 1026-1030.

[3] R. Garcia-Guzman, Y. A. Andrade-Ambriz, M.-A. Ibarra-Manzano, S. Ledesma, J. Carlos Gomez, and D.-L. Almanza-Ojeda, Trend-based categories recommendations and age-gender prediction for pinterest and twitter users, Applied Sciences. 17 (2020), 1-12.

[4] Z. Xu, L. Chen, Y. Dai, and G. Chen, A dynamic topic model and matrix factorization-based travel recommendation method exploiting ubiquitous data, IEEE Transactions on Multimedia. 8 (2017), 1933-1945.

[5] http://www.openu.ac.il/home/hassner/Adience/data.html.

[6] G. Levi, T. Hassner, Age and gender classification using convolutional neural networks, In Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015, 34-42.

[7] H. Meinedo, I. Trancoso, Age and gender classification using fusion of acoustic and prosodic features, In Eleventh annual conference of the international speech communication association. 2010.

[8] A. Ekmekji, Convolutional neural networks for age and gender classification, Stanford University. 2016.

[9] M. Duan, K. Li, C. Yang, and K. Li, A hybrid deep learning CNN–ELM for age and gender classification, Neurocomputing. 275 (2018), 448-461.

[10] S. Lapuschkin, A. Binder, K.-R. Muller, and W. Samek, Understanding and comparing deep neural networks for age and gender classification, In Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017, 1629-1638.

[11] J. Tang, S. Wang, X. Hu, D. Yin, Y. Bi, Y. Chang, and H. Liu, Recommendation with social dimensions, In Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[12] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, Recommender systems with social regularization, In Proceedings of the fourth ACM international conference on Web search and data mining. 2011, 287-296.

[13] T. Ameen, L. Chen, Z. Xu, D. Lyu, and H. Shi, A convolutional neural network and matrix factorization-based travel location recommendation method using community-contributed geotagged photos, ISPRS International Journal of Geo-Information. 8 (2020), 1-16.

[14] A. Mnih, R. R. Salakhutdinov, Probabilistic matrix factorization, In Advances in neural information processing systems. 2008, 1257-1264.

[15] S. Kisilevich, F. Mansmann, and D. Keim, P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos, In Proceedings of the 1st international conference and exhibition on computing for geospatial research & application. 2010, 1-4.

[16] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv. 2014, 1-14.

[17] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, Author topic model-based collaborative filtering for personalized POI recommendations, IEEE transactions on multimedia. 6 (2015), 907-918.

[18] X. Liu, X. Qian, D. Lu, X. Hou, and L. Wang, Personalized tag recommendation for Flickr users, In IEEE International Conference on Multimedia and Expo (ICME). 2014, 1-6.

[19] D. Zhou, B. Wang, S. M. Rahimi, and X. Wang, A study of recommending locations on location-based social network by collaborative filtering, In Canadian Conference on Artificial Intelligence. 2012, 255-266.

[20] H. Gao, J. Tang, X. Hu, and H. Liu, Exploring temporal effects for location recommendation on location-based social networks, In Proceedings of the 7th ACM conference on Recommender systems. 2013, 93-100.

[21] S. Yadav, S. Yadav, A. Bhangale, and A. Bobade, A Geo-PFM Model for Point Of Interest Recommendation. 2017, 1534-1536.

[22] Y. Hu, Y. Koren, and C. Volinsky, Collaborative filtering for implicit feedback datasets, In IEEE International Conference on Data Mining. 2008, 263-272.

[23] S. Wang, Y. Wang, J. Tang, K. Shu, S. Ranganath, and H. Liu, What your images reveal: Exploiting visual contents for point-of-interest recommendation, In Proceedings of the 26th international conference on world wide web, 2017.