

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

1ST WAY

pdf to data extraction

```
In [9]: import PyPDF2
reader= PyPDF2.PdfReader('Final nm.pdf')
print("total pages", len(reader.pages))

#extract text
page= reader.pages[0]
#print(page.extract_text())

#whole text data extract
for i in range(len(reader.pages)):
    page= reader.pages[i]
    print(page.extract_text())
```

total pages 15

See discussions, stats, and author profiles for this publication at : <https://www.researchgate.net/publication/335060973>

User Transition Pattern Analysis for Travel Route Recommendation

Article in IEICE Transactions on Information and Systems · August 2019

DOI: 10.1587/transinf.2019EDP7096

CITATIONS

7READS

325

3 authors:

Junjie Sun

Kyoto University

7 PUBLICATIONS 22 CITATIONS

SEE PROFILE

Chenyi Zhuang

Kyoto University

21 PUBLICATIONS 598 CITATIONS

SEE PROFILE

Qiang Ma

Kyoto Institute of Technology

130 PUBLICATIONS 1,053 CITATIONS

SEE PROFILE

All content following this page was uploaded by Junjie Sun on 04 December 2019.

The user has requested enhancement of the downloaded file.

VOL. E102-D NO. 12

DECEMBER 2019

The usage of this PDF file must comply with the IEICE Provisions on Copyright.

The author(s) can distribute this PDF file for research and educational (nonprofit) purposes only.

Distribution by anyone other than the author(s) is prohibited.

2472IEICE TRANS. INF. & SYST., VOL.E102-D, NO.12 DECEMBER 2019

PAPER

User Transition Pattern Analysis for Travel Route Recommendation

Junjie SUN†), Chenyi ZHUANG††, Nonmembers, and Qiang MA†, Member

SUMMARY A travel route recommendation service that recommends

a sequence of points of interest for tourists traveling in an unfamiliar city is a very useful tool in the field of location-based social networks. Although there are many web services and mobile applications that can help tourists to plan their trips by providing information about sightseeing attractions, travel route recommendation services are still not widely applied. One reason could be that most of the previous studies that addressed this task were based on the orienteering problem model, which mainly focuses on the estimation of a user–location relation (for example, a user preference). This assumes that a user receives a reward by visiting a point of interest and the travel route is recommended by maximizing the total rewards from visiting those locations. However, a location–location relation, which we introduce as a transition pattern in this paper, implies useful information such as visiting order and can help to improve the quality of travel route recommendations. To this end, we propose a travel route recommendation method by combining location and transition knowledge, which assigns rewards for both locations and transitions. key words: travel route recommendation, sightseeing, location-based so-

utions. To this end, we propose a travel route recommendation method by combining location and transition knowledge, which assigns rewards for both locations and transitions. key words: travel route recommendation, sightseeing, location-based so-

cial network, matrix factorization

1. Introduction

In recent years, tourism has become one of the most important industries in the world. At the same time, benefiting from the rapid development of location-based social networks (LBSN), user-generated content (for example, photos, check-ins, and ratings) is increasingly available on the Internet. Various services exist to help tourists with a better tourism experience when they travel in an unfamiliar urban area (for example, point of interest (POI) recommendation). One of the important tasks is travel route recommendation,

or tour planning, which aims to automatically recommend travel routes that meet users' requirements; this is addressed in this paper. Currently, most existing research on, and services for, travel route recommendation are based on the model of the orienteering problem (OP) and its variants [1]. OP is a classical route planning problem whose objective is to maximize the total score on a path passing through a subset of nodes without exceeding the time budget, on an undirected

graph with weighted nodes and known travel time between these nodes [2]. For the travel route recommendation task, an urban area is a graph, a POI is a node in the graph, the

Manuscript received April 3, 2019.

Manuscript publicized September 6, 2019.

†The authors are with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606–8501 Japan.

††The author is with the National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, 135–0064 Japan.

a) E-mail: jj-sun@db.soc.i.kyoto-u.ac.jp

DOI: 10.1587/transinf.2019EDP7096 time cost to travel from one POI to another is the weight of

an edge, and users' satisfaction is represented as the node weight (that is, tourists visit a POI to collect its score). The output is an optimal travel route that maximizes the total user satisfaction score while keeping the total travel time

within the budget.

[3] described three key relations in LBSN studies:

(1) user–user relations, (2) user–location relations, and

(3) location–location relations. Conventional approaches

mainly include node rewards (that is, user–location relations) in their objective functions, while the relations between locations are considered as the travel distance in constraints. In other words, their objective function assumes that tourists collect rewards after visiting locations and find

the route with the maximal total reward. However, these methods suffer from several drawbacks, listed below:

- Budget–reward trade-off. Because the objective is to

maximize the route's reward, the recommended route will be a trade-off between the travel budget and lo-

cation rewards, subject to several constraints. Taking the travel time budget constraint as an example, to

maximize the route's reward, conventional approaches

will avoid locations requiring long travel duration and add locations with a high reward-time ratio. However, such locations may not be of high sightseeing quality (because locations with shorter travel duration have a higher reward-time ratio), so some landmarks with a

long travel duration might be ignored during the tourplanning.

•Insufficient consideration of location–location relations. Failing to consider location–location relations might reduce the recommendation quality. For in-stance, various influences (for example, public transportation, geographical constraints, and travel guides) may cause tourists to always visit some POIs in a certain order. Disregarding the visiting order and always recommending the shortest route may generate unreasonable routes. Figure 1 shows an example of different visiting orders. Each dimension of the figure represents locations in the Toronto dataset (introduced in Sect. 5.1) and each entry represents the normalized observed transition weight from one POI to another. We observe that transitions in the green rectangle are asymmetric, whereas they are symmetric in the red rectangle; the green rectangle demonstrates the unbalanced visiting orders between locations.

Copyright © 2019 The Institute of Electronics, Information and Communication Engineers

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2473

Fig. 1 Visualization of the transition matrix on the Toronto dataset. Each dimension represents locations in the city.

Recently, several studies have attempted to improve the quality of the travel route recommendation system by adjusting the OP-based model. For example, based on the OP, [4] studies the unsatisfactory recommendations caused by

the scope of the POI. The authors illustrate that some POIs are large areas and not suitable for representation as a single POI, and propose a method to jointly consider outer and inner travel routes across different ranges of POIs.

To address the above drawbacks and improve the recommendation quality, in this paper, we propose a travel route recommendation method by introducing user transition patterns into the OP model. In general, a travel route consists of nodes (locations) and edges (transitions between

locations). From the travel routes of many tourists, we can not only discover popular visiting locations but also transition patterns between locations. As tourists travel from one location to another in the city, they make different transitions

with different frequencies (weights); we introduce these as transition patterns in this paper. An idea similar to this paper is considered in [5]. Both location and transition knowledge are jointly modeled to recommend a travel route. They take the travel route recommendation task as a likelihood maximization problem and propose a method to compute

probabilities of points and transitions. However, the joint model has been evaluated only under the route length budget and does not show the advantages of introducing transition knowledge.

Transition patterns can be regarded as connections between locations, whereby the higher the weight of a transition, the tighter the relationship between the two locations. A transition pattern implies not only information such as visiting order but also possibly a scenic path between locations

(for example, featured streets between POIs or beautiful paths in a large area). In this way, the location–location relations are

emphasized.

A large amount of users' transition data helps us to generate a knowledge graph. This represents the target travel city's implicit structure information, such as visiting order and tourist flow trends, and helps to restrict the route to be matched with the real situation. For instance, Fig. 2 shows part of the visualized observed transition weight graph of Edinburgh in the dataset (introduced in Sect. 5.1). The

Fig. 2 Visualization of the observed transition weight graph in Edinburgh; a train station is located in the yellow rectangle. The thickness of each edge represents the weight (transition count).

Thickness of an edge represents the weight of the transition (that is, how many tourists traveled through the edge in the dataset). Basically, POIs in the center of the city are tightly connected to each other and loosely connected to the outercity. In the yellow rectangle, there is a train station. This can be regarded as a type of geographical constraint because most transitions pass on both sides of the station but fewer

pass by the station. This type of restriction should be considered in the travel route recommendation.

Therefore, we extract transition patterns from users' travel route data and use them to compute rewards on edges. This changes the OP to a mixed orienteering prob-

lem (MOP) [6]: both nodes and edges are assigned rewards.

We recommend travel routes by solving this MOP, aiming to maximize the total reward from both nodes and edges.

Our main contributions are summarized as follows:

- We propose a general framework for recommending travel routes under different travel budget constraints. The experimental results verified that our method, which considers both locations and transitions, can achieve better performance than recently proposed methods.
- We propose a novel method to learn transition patterns by incorporating them with spatiotemporal features. Experimental results reveal the effectiveness of the transitions learned by introducing both spatial and temporal influences.
- A parameter study is included, to understand the balance between location and transition in the tour planning procedure.
- A real tourism dataset of Kyoto is applied for real user evaluation. We compare the recommendation results with and without transition knowledge, which confirms our hypothesis that transition patterns are helpful for improving the quality of travel route recommendations.

2. Related Work

2.1 Tourist Trip Design Problem

The tourist trip design problem (TTDP) is widely studied in operational research, and OP is suitable as a theoretical model of it. OP-variant models that satisfy more realistic constraints have been proposed. These studies have focused on proposing suitable heuristic planning algorithms,

2474 IEICE TRANS. INF. & SYST., VOL. E102-D, NO. 12 DECEMBER 2019

because it is an NP-hard task. For example, team OP (TOP) extends OP by allowing multiple tours, and is applied for multi-day tour planning (e.g., [7]). Unlike the TOP, a recent study [8] proposes a multi-day tour planning algorithm by maximizing the utility of the worst day.

The most studied extension of OP-based tour planning is TOP with time windows (TOPTW). This adds opening and closing times for each node (that is, POIs' opening hours) to TOP, and was solved with an iterated local search (ILS) algorithm in [9]. Based on this extension, public transportation services were integrated into [10]. [11] is probably the most complete study of tour planning based on TOPTW; the authors jointly modeled various settings mentioned in previous studies (for example, public transportation service and opening hours). In [4], the authors defined super-POI

for the large sightseeing areas, which contain smaller sightseeing spots, and separately plan the outer and inner routes. However, none of these works has considered modeling users' travel behavior or directly planning the travel route with predefined values (for example, POI reward).

2.2 Tour Recommendation There are many tour recommendation methods that have been proposed, based on the OP model, to provide a better recommendation quality. These are more concerned with how best to discover and assign reward values for POIs in the OP model. For instance, personalization has been ad-

dressed by considering users' travel behavior and assigning more weight to users' preferred POIs.

One of the earliest OP-based tour recommendation methods, proposed in [12], discovered POIs by mining from a large-scale dataset of geotagged social images. It has been extended by Lim et al. [13], to mine social geotagged images' metadata and estimate users' preferences with time-based interests. A similar idea was considered in [14], which estimates users' preferences by the categories of the POIs. With the development of computer vision technology, an image feature-enhanced tour recommendation method was proposed in [15], to find POIs that are visually similar to users' uploaded images. Preferences can be manually selected in [16], [17]; the authors developed an interactive web system in which users can manually decide travel preferences. A feature-centric matrix factorization model was applied in [18]. With a user-feature matrix, preferences that out of the residential areas also can be evaluated through the collaborative filtering model. [5] is the study most related to our work: it recommends travel routes based on both POIs and transition probabilities, within a travel length budget.

There have also been several studies not based on the OP model. For instance, [19] models the tour planning task as a max-cover problem that finds the most suitable travel routes from other tourists' travel history, according to key-word matching. However, the literature reviewed above rarely considered location-location relations, and recommended travel routes based only on user-location matching. In our work, we study location-location relations by considering transition patterns to improve the existing tour recommendation method. Different from our previous work [20], we modified the transition pattern inference method and added more experiments and evaluations.

3. Preliminaries

3.1 Problem Definition

We first define two concepts: travel route and user query, and then define our travel route recommendation task.

Definition 1 (travel route): A travel route is a sequence of

POIs (p_1, p_2, \dots, p_L). Each point p is a location that the tourist has visited, and consists of {route identifier, POI identifier, category, date and time, longitude, and latitude }.
Definition 2 (user query): A user query is a query $q = (p_s, p_e, B)$ in which p_s and p_e represent the start and end point, respectively, and B represents the travel budget. We use Q to denote all of the queries corresponding to the travel route data. In our work, we consider two types of travel budgets: the route length budget B_l and the time budget B_t .

With these two concepts, our problem is defined as follows:

Problem 1: Assuming there are N POIs, $P = \{p_1, p_2, \dots, p_N\}$, and tourist travel route data for the target city, given a user query, the framework will output a list of POIs of P as a recommended travel route.

3.2 Objective Function

According to the problem definition, we assume that a user can receive a reward by visiting a POI or traveling between POIs. The reward here represents a user's satisfaction with this tour. Therefore, we recommend a travel route by maximizing the total reward of the tour, which includes the rewards on both locations and transitions. Specifically, given N POIs, P , and tourists' travel route data in the travel city, a travel route is recommended according to the user query by solving the following objective function [13]:

$$\begin{aligned}
 & \max \sum_{i=1}^{N-1} \sum_{j=2}^N R(p_j | p_i) x_{ij} \quad (1) \\
 & \text{s.t.} \sum_{j=2}^N x_{ij} = 1, \forall i = 1, 2, \dots, N-1 \quad (2) \\
 & \sum_{i=1}^{N-1} x_{ij} = 1, \forall j = 2, 3, \dots, N \quad (3) \\
 & \sum_{i=1}^{N-1} \sum_{j=2}^N x_{ij} = B_l - 1 \quad (4) \\
 & p_i - p_j + 1 \leq (N-1)(1 - x_{ij}), \forall i, j = 2, \dots, N \quad (5)
 \end{aligned}$$

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2475

where $R(p_j | p_i)$ is the reward function for visiting p_j from p_i , which will be defined in the following sections; N is the available number of POIs in the travel city; x_{ij} is a binary indicator that equals 1 when a user travels from POI p_i to p_j and equals 0 otherwise. Here we mark p_1 as the start location and p_N as the end location. Constraint 2 specifies that a route must begin at the start location and finish at the end location. It also prevents a travel route from revisiting the start location or traveling from the end location. Constraint 3 specifies that each POI can only be visited once.

Constraint 4 is the travel budget constraint, which restricts

the recommended travel route to be no longer than B
 $l-1$

(B is the number of POIs that the tourist wants to visit).

Constraint 5 specifies that subtours are to be avoided; this constraint was proposed in the classical traveling salesman problem [21].

In addition, we consider another budget constraint—

the travel time budget—which limits the recommended route's total travel time to be no greater than B

t. The following

function defines the time budget constraint and replaces constraint 4:

$$\sum_{i=1}^{N-1} \text{Cost}(p_i, p_{j_i}) \leq B_t, \quad (6)$$

where $\text{Cost}(p_i, p_j)$ includes the travel time from p_i to p_j and

the visiting duration time at p_j ; B_t represents the total travel time budget.

The objective function is similar to other OP-based methods that aim to maximize tourists' visiting rewards.

The key point is how to correctly define the reward function $R(p_j | p_i)$. We introduce different methods to assign rewards to locations and transitions; the reward function $R(p_j | p_i)$ is then defined as a combination of these two types of rewards.

4. Reward Functions 4.1 Location Reward Location reward represents how satisfied the tourist is after

visiting the location. Most of the related literature focuses on this, by considering the user–location relation. Because it has been widely studied, we directly apply existing work

to assign rewards to locations. Specifically, we adopt two different methods, to adapt to different travel budget constraints and to test the generality of our proposed framework.

POI ranking. A naive idea to assign a reward to a location

is by using POI popularity, which assumes that the more tourists visit a location, the more attractive it is. As an alternative, machine learning methods allow us to take advantage of more features than POI popularity, to train a

model with other users' travel route data. For instance, the

rank support vector machine (RankSVM) is applied in [5],

by introducing more features (e.g., category, popularity, and average visit duration) to rank POIs according to the user query under the travel length budget constraint. Then the POI ranking scores can be obtained as POI reward by learning

the ranking of POIs using RankSVM, with a linear kernel and L2 loss [22]. The objective function is:

$$\min$$

$$\|w\|$$

$$2w^T w + C \sum_{p_i, p_j \in P, q \in Q} \max(0, 1 - w^T (\phi(p_i, q) - \phi(p_j, q)))^2, \quad (7)$$

where w is the parameter vector, $C > 0$ is a regularization

constant, P is the set of POIs to rank, and Q represents all of the queries with respect to the travel routes in the training set; ϕ is the

feature vector for a POI p with respect to a query q .

Given a query q , the ranking score of POI p is then

computed as:

$$R_{p,j,q} = wT_{p,j,q} \quad (8)$$

All ranking scores are scaled to the range [0, 1] with the following softmax function:

$$P(p_j | p_i) = \frac{\exp(R_{p,j,q})}{\sum_{p_j} \exp(R_{p,j,q})} \quad (9)$$

where $R_{p,j,q}$ and $R_{p,x,q}$ are ranking scores computed with

$$\text{Eq. (8)}.$$

Time-based User Interest. In addition, we consider

the factor of personalization: that tourists may be more interested in their preferred locations. Because we cannot

manually decide user preference for queries in Q , we adopt

time-based user interest [13] to automatically estimate personalized preference, in different POI categories. This is

based on the heuristic idea that tourists would stay longer when visiting the category of POIs they like. Time-based

user interest is denoted as $IntTime$

$u(Cat_{p_j})$ which can be computed by comparing each user's duration on certain POI category to the average visit duration of all users.

With the computed user interest, the personalized location reward can be decided as follows:

R

$$P(p_j | p_i) = \eta IntTime$$

$$u(Cat_{p_j}) + (1 - \eta) Pop(p_j), \quad (10)$$

where η is the trade-off parameter that decides the weight

between personalized interest $IntTime$

$u(Cat_{p_j})$ and POI's popularity $Pop(p_j)$. Specifically, we set $\eta = 0.5$ as described

in [13]. We then normalize each POI reward value by the

maximum value, to scale all values to [0, 1].

4.2 Transition Reward To address the drawbacks of OP-based methods and improve the quality of the recommended travel route, we

would like to take location-location relations into consideration. Specifically, transition patterns can be regarded as the weights of going from one POI to another, which contains certain features such as visiting order. We extract transition

patterns from tourist travel route data and assign a reward to

each edge between two POIs.

One naive idea to represent transition pattern is to directly normalize the observed transitions. Then the transition patterns between POIs could be regarded as the rewards

of transitions. However, this idea has several drawbacks:

• The transition data that we can observe are not complete; that is, some locations that have transitions may not be observed in our datasets.

• Such a model cannot deal with new POIs; that is, there are no transitions that can be observed when new POIs are discovered and added to the dataset.

Therefore, given the travel route of all tourists, we extract the number of transitions (user transits from one location to another) for each pair of POIs. Then an observed transition matrix T can be constructed, where N is

the number of POIs in the target travel city and each entry in T denotes the transition frequency times between POIs. Because some of the transitions are not observed, we aim to infer the completed transition matrix \hat{T} by exploiting observed transitions to predict transition values to un-observed entries. Specifically, we divide each entry by the maximum entry value to normalize \hat{T} and assign transition reward as below:

$$R(p_j | p_i) = \hat{T}_{i,j} / \max_j \hat{T}_{i,j} \quad (11)$$

In the following subsections, we propose a spatiotemporal feature enhanced matrix factorization model to assign transition rewards.

Weighted Transition Matrix Factorization. Similar to the idea of collaborative filtering, POIs may have common attractions or connections to other POIs, owing to common features that match very well. For instance, a popular landmark has a stronger connection to other locations than some unpopular POIs because many transits are observed that travel to or from it. A common technology for collaborative filtering in the field of recommender system research [23] is matrix factorization, which factorizes the interaction matrix into two lower-dimensional latent feature matrices. An unobserved interaction entry can then be predicted through the inner product of two latent feature matrices. Therefore, we can infer unobserved transition weights

by exploiting matrix factorization technology.

Unlike related work that applied matrix factorization for user-item or user-location interaction pairs [23], [24], we want to use it to study location-location relations (that is, both decomposed lower latent features are about locations). We first construct the observed transition matrix T

by extracting transitions from tourist travel route data. We then divide each entry by the maximum entry value to normalize T and obtain a matrix T , in which the value of each entry is in $[0, 1]$ and represents the relative weight of each transition.

One naive solution would be to directly factorize the weighted transition matrix. However, because the features of the two dimensions in the matrix are the same (both about locations), we add an interaction latent component M to represent local interaction models of the decomposed latent variables [25]. Hence, one reasonable solution for inferring the weighted transition matrix from observed data is to factorize the observed weighted transition matrix T as follows:

$$T \approx V_s M V_t^T \quad (12)$$

where $V_s \in \mathbb{R}^{N \times k}$ and $V_t \in \mathbb{R}^{N \times k}$ represent latent features of source and destination points, respectively; $M \in \mathbb{R}^{k \times k}$ is the interaction matrix that represents the relationship between locations; and k is the number of latent features. The latent matrices V_s , V_t , and M are then computed by solving the following optimization function:

$$\min_{V_s, V_t, M} \|T - V_s M V_t^T\|_F^2$$

$$t)/\text{vextenddouble}/\text{vextenddouble}/\text{vextenddouble}2+\lambda r/\text{bracketleftBig} \\ / \text{bardblVs}/\text{bardbl2+}/\text{bardblM}/\text{bardbl2+}/\text{bardblVt}/\text{bardbl2}/\text{bracketrightBig}$$

(13)

where $\|\cdot\|_2$ denotes the Frobenius norm; I is a binary matrix each entry of which, i, t , indicates whether a transition has been observed; and λ is the parameter for the regularization term.

With learned latent factors V_s, V_t , and M , the unobserved transition entries are filled with predicted values and an inferred weighted transition matrix \hat{T} can be computed through Eq. (12). Finally, we assign the transition reward $R(T(p_j|p_i))$ through Eq. (11).

Spatial Influence. Although we can infer a complete weighted transition matrix by directly factorizing Eq. (13), there is still scope for improving the performance. For instance, spatial influence is an important factor in the relation of a pair of POIs. [24], [26] introduced spatial influence as

the physical distance between user and location or between user and user. In contrast, we consider spatial influence as a feature of the relation between locations, which can be applied to enhance our weighted transition matrix factorization model.

Figure 3 shows the cumulative distribution function of transition distance on two different datasets: Glasgow and Toronto. We can see that both curves rise dramatically when the distance is small: most tourists travel at most 2 km from one POI to the next. Tourists tend to visit nearby POIs; in other words, closer POIs have a stronger influence. Therefore, we want to extract location–location spatial features by considering the distance between POIs.

First, we calculate the distance between POIs using the haversine formula [27] (to compute the distance between two geographic coordinates). Second, we take the reciprocal of the value in each entry to capture a POI distance reciprocal matrix, in which the longer the distance, the smaller the value. Finally, we normalize each entry by dividing by the

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2477

maximum value in the matrix, to construct the POI spatial influence matrix $G \in \mathbb{R}^{N \times N}$. The spatial influence matrix can be regarded as additional global knowledge under the assumption that the closer pairs of POIs are more likely to be visited together. In other words, each entry represents

the confidence of location–location spatial influence, which is larger when the distance is smaller.

We decompose G to get the lower latent spatial features of location V

g. Unlike Eq. (12), in which source and destination represent different latent variables, the distance between two locations is the same in both directions, so we use the same latent variable $V_g \in \mathbb{R}^{N \times k}$ to represent the spatial feature of location. Therefore, the spatial influence matrix can be factorized as follows:

$$G \approx V$$

$$gV^T$$

g, (14)

We can obtain the latent spatial feature V_g by solving the objective function as follows (also known as nonnegative matrix factorization):

min

$\|V_g - V_g V_T^T\|_F^2 + \lambda \|V_g\|_F^2$ (15)

where λ is the parameter for the regularization term.

Temporal Influence. Temporal influence is another

important factor for LBSN studies. For example, users' check-in time flow was considered in [28], which models

temporal influence as category preference over time (for example, visiting the office in the morning or the gym in the evening). However, because our target is to enhance the representative ability of location–location relations, we focus

more on the temporal influence of POIs. Similar to spatial influence, we aim to obtain location–location temporal features to enhance the latent features V

and V_t .

There are many temporal factors related to POIs, users, or travel routes. We consider several of them that are helpful to obtain location–location temporal features. In terms of

spatial influence, we assume that a pair of POIs with smaller distance has a larger transition weight. However, this might not work well for all POIs. For

instance, users may prefer to

go to a restaurant at noon and evening. Suppose a user visits

a POI near a restaurant in the morning. From the viewpoint of spatial influence, the restaurant could be selected to visit next because of its higher transition weight. However, from the viewpoint of temporal influence, the restaurant

would not be selected because the time is inappropriate to visit.

Therefore, by investigating tourist travel route data, we have two observations:

- Each POI has an available open time window and visits vary over 24 hours. As shown in Fig. 4 (a), the peak times for visiting are in the morning and afternoon, whereas the peak in Fig. 4 (b) is around noon.
- For every travel route of users, Fig. 5 shows the distribution of intervals between users' visiting times for two POIs. These indicate how long tourists prefer to travel from one POI to the next: most of the transits are

Fig. 4 Visiting time distributions. Examples of two POIs over 24 hours.

Fig. 5 Visiting time interval distributions at two cities. under two hours.

The different visiting times in Fig. 4 may be caused by different POI categories and users' transitions between POIs. We can make the simple assumption that the transition weight between two POIs is greater when the difference between their visiting time peaks is more suited to user transit behavior (that is, smaller than a certain value of visiting time interval). We aim to extract latent spatial features of POIs, which helps to restrict the influence weight upon the spatial influence features.

With the above observations and assumptions, we use a Poisson distribution to fit users' visiting time intervals and a Gaussian mixture model (GMM) to fit the check-in time data of each POI. We a

automatically fit the data and set the GMM component range from 1 to 3, which corresponds to morning, afternoon, and night. We then construct the temporal influence matrix E

by checking the difference

in the means of visiting time distribution between two POIs and taking this value into the fitted visiting time interval distribution to fill the entry value of E

where E_{ij} is equal to

the normalized E_{ij} , which represents the temporal influence between POIs. We represent this temporal influence matrix

as below:

$E \approx V_e V_t^T$

(16)

where V_e represents the latent temporal features of POIs. To

reduce the computational complexity, we assume that the latent temporal features for source and destination are the

same. Similar to the spatial influence matrix G , we obtain

the latent spatial feature V

by solving the following objective function:

min

$\|V_e - V_e V_t^T\|_F^2 + \lambda \|V_e\|_F^2$

(17)

2478 IEICE TRANS. INF. & SYST., VOL. E102-D, NO. 12 DECEMBER 2019

Fig. 6 Visualization comparison of weighted transition matrix and distance reciprocal matrix on Edinburgh dataset.

where λ is the parameter for the regularization term.

Having obtained the latent spatial and temporal features,

we face the question of how to introduce them into

the weighted transition matrix factorization model to enhance the inference ability. By considering the above observations of user transition behaviors, we want the latent

location features to be more similar to the latent spatial and

temporal features. Figure 6, which visualizes the correlation between the latent location and spatial features, supports this viewpoint. For both Figs. 6 (a) and 6 (b), each dimension represents locations in the Edinburgh dataset.

Figure 6 (a) shows the observed weighted transition matrix

while Fig. 6 (b) shows the spatial influence matrix (the distance reciprocal matrix). The larger the value, the darker the color. The visualization shows that they are quite similar to each other, which verifies the close relationship between the

latent location features and latent spatial features.

Therefore, we introduce the spatial and temporal influences

to enhance Eq. (12) by minimizing the difference

between the latent location and spatiotemporal features. In

other words, we not only require the latent location features

V

and V_t to produce less error, but also want them to be as

similar as possible to V_g and V_e . Specifically, we use the

following objective function to factorize the weighted transition matrix:

min

$\|V_s, V_t, M - V_s V_t^T\|_F^2 + \lambda \|V_s\|_F^2 + \lambda \|V_t\|_F^2$

(18)

$$\begin{aligned}
& e/v_{\text{extenddouble2}}/bracketrightBig \\
& +\lambda e/bracketleftBig \\
& /bardblVs-V_e/bardbl2+ /bardblVt-V_e/bardbl2/bracketrightBig \\
& +\lambda r/bardblM/bardbl2, \\
& (18)
\end{aligned}$$

where λr is the parameter for the regularization term, and λg and λe control the weights of the spatial and temporal influence, respectively. For instance, with a large value on λg , the latent location features are more similar to the latent spatial features V_g .

Finally, the inferred weighted transition matrix can be computed with Eq. (12) and the transition reward $RT(p_j|p_i)$ assigned through Eq. (11).

4.3 Combining Location and Transition Reward After assigning rewards to locations and transitions, we now aim to combine knowledge of locations and transitions to recommend travel routes. Both location and transition rewards are already normalized and each value is in $[0, 1]$. Therefore, we can combine the location and transition rewards with the following equation:

$$R(p_j|p_i) = \alpha RP(p_j) + (1-\alpha)RT(p_j|p_i), \quad (19)$$

where $R(p_j|p_i)$ is the reward function defined in the objective function Eq. (1); $RP(p_j)$ and $RT(p_j|p_i)$ are the location and transition reward functions defined in Sects. 4.1 and 4.2, respectively; $\alpha \in (0, 1)$ is a trade-off parameter that indicates the relative importance of location and transition rewards.

We briefly review our definition of reward function, which considers location, transition, and their combination. If we use location reward, the objective function Eq. (1) is similar to the original OP [2], which only includes reward on

nodes. If we use transition reward, it is similar to the arc orienteering problem (AOP) [29], which only includes reward on edges. In terms of the combination equation Eq. (19), the reward function $R(p_j|p_i)$ now includes both nodes and

edges of the POI graph, which leads the objective function Eq. (1) to be a mixed orienteering problem (MOP) [6]. Although this is quite different from the original OP, it can still be treated as a mixed integer linear problem, which can be solved by optimization tools such as Gurobi for lp solve++.

4.4 Optimization and Latent Variable Learning

We solve our travel route recommendation objective function Eq. (1)–(5) in Sect. 3.2 with the Gurobi optimization package. For the training strategy, we minimize the objective functions Eq. (13), (15), (17), and (18) with a gradient descent approach by iteratively optimizing the latent variables V_s, V_t, M, V_g , and V_e ; this is supported by the Theano framework.

5. Experiments

5.1 Experimental Setup

To evaluate our proposed method, both quantitative and qualitative experiments were conducted.

Datasets. To evaluate the recommendation performance of our proposed framework, a public LBSN dataset

was used [5],[13]. This dataset contains travel route data extracted from Flickr photos in five cities: Budapest, Edinburgh, Glasgow, Toronto, and Vienna. The detailed statistics of the dataset are shown in Table 1. In our travel route recommendation experiments, we randomly divided the data of each city into five folds, and performed a five-fold cross-validation [30] to evaluate different approaches. This means that, when testing on one fold of the dataset, we used the other folds of data to train our models.

To evaluate our proposed method with real tourists, we

†Gurobi Optimization. <http://www.gurobi.com>

‡lpsolve package: <http://lpsolve.sourceforge.net>

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2479

Table 1 Statistics of the Flickr image dataset.

Dataset	#Photos	#Check-ins	#Travel Routes	#Users
Budapest	145,364	18,513	2,361	954
Edinburgh	82,060	33,944	5,028	1,454
Glasgow	29,019	11,434	2,227	601
Toronto	157,505	39,419	6,057	1,395
Vienna	461,905	1,155	34,515	3,193

adopted a real tourism dataset of Kyoto. This dataset orig-

inally contained GPS trail data for 450 foreign tourists and 406 students on school excursion one-day tours. It records

the GPS signal every two seconds for each tourist, so a travel

route can be formed by linking all these GPS points.

According to the GPS trails, we first mapped all GPS

points to 123 POIs that represent sightseeing attractions. We used these data to recommend travel route for real tourists.

Queries. To evaluate our proposed framework's gen-

erality, we considered two types of common travel budget (route length budget and travel time budget), which were introduced for tour recommendation in

[5],[13]. For simplic-

ity, we use a traveling speed of 4 km/h (a leisurely walking

speed) as a tourist's average speed when traveling between

POIs under the travel time budget; this has also been applied for tour recommendation by [13],[15].

For both travel budget constraints, we use queries ex-

tracted from tourists' real travel route sequences. Travel

route sequences shorter than three POIs are ignored in the evaluation because they cannot be presented as valid queries. For instance, suppose a tourist started from p

3 and

traveled to p6, visiting three POIs and spending four hours

on this trip. Then a query can be constructed as $q(p_3, p_6, 5)$

and $q(p_3, p_6, 4)$, for route length budget and travel time budget, respectively.

For real tourist evaluation, two queries are tested:

- $q_1 = (\text{Kyoto National Museum}, \text{Sanjo}, 4h)$

- $q_2 = (\text{Kyoto University}, \text{Sanjo}, 7h)$

These two queries consider different start locations and

travel time budgets, finally returning to the hotel near Sanjo. We take the average visiting duration as the travel time for each POI and the travel speed between POIs is set to an av-

erage speed of 12 km/h for simplicity (this can be easily re-

placed by other transportation data: for example, transport time computed by Google maps).

5.1.1 Comparison Methods

To understand the effectiveness of our proposed framework, we compared it with a list of approaches. Under the route length budget, the following approaches were tested:

- PoiPop. A baseline approach that recommends travel route based only on POI popularity.
 - PoiRank. A machine learning method proposed in [5]. By training a RankSVM model with POI features under the travel length budget, it recommends a travel route based on the POI ranking score.
 - Markov. An explicit feature factorization method which factorizes a pair of POIs by different types of POI features to compute the transition probabilities between POIs. It is similar to our defined transition reward except that the sum of the outer links is equal to 1.
 - Rank + Markov. A combination method, proposed in [5], that jointly optimizes point preferences and transition probabilities.
 - Tmf, GTmf, TGTmf. Our proposed methods, introduced in Sect. 4.2, that recommend travel routes based only on transition reward. Tmf directly factorizes the weighted transition matrix. Spatial influence features are introduced in GTmf. Spatial and temporal influences are jointly modeled in TGTmf.
 - Rank + Tmf, Rank + GTmf, Rank + TGTmf. Our proposed combination methods in Sect. 4.3 to recommend travel routes based on both location and transition rewards. Location reward is assigned with PoiRank and transition reward is assigned with our proposed Tmf, GTmf, and TGTmf, respectively.
- Under the travel time budget, we combine PersTour in Sect. 4.1 with our proposed Tmf, GTmf, and TGTmf, described above, to evaluate the performance of travel route recommendation under the travel time budget.
- PersTour. Proposed in [13], which recommends travel routes by estimating personalized time-based user interest from users' travel history data.
 - PersTour + Tmf, PersTour + GTmf, PersTour + TGTmf. Our proposed combination methods that jointly consider location and transition rewards. PersTour is applied to assign location reward and transition reward is given by Tmf, GTmf, and TGTmf, respectively.

5.2 Evaluation Metrics

Various quantitative evaluation metrics, introduced in related work for travel route recommendation, are applied to our evaluation. Evaluation on Travel Route Recommendation. One common evaluation in recommender system research is to compare the recommended result with the user's selection in real life. Therefore, we evaluate all of the methods by comparing the recommended travel route with the user's actual travel route in query set Q . The performance of all approaches is evaluated by using the following metrics.

- Tour F

1score: TF1. Because the F1 score is a common metric for evaluating POI and travel route recommendations ([5],[13],[31]), we use the tour F

lscore as
an evaluation metric for the recommended travel route.
The tour F lscore is the harmonic mean of the tour recall (how many of the user's real visited POIs are recommended) and tour precision TP (how many recommended POIs are in the user's real travel route). It is applied to a recommended travel route excluding the

2480 IEICE TRANS. INF. & SYST., VOL. E102-D, NO. 12 DECEMBER 2019

Table 2 Performance comparison of travel route recommendation under the route length budget in terms of tour F lscore (TF1).

Budapest Edinburgh Glasgow Toronto Vienna

PoiPop 0.216±0.277 0.331±0.378 0.346±0.429 0.200±0.320 0.316±0.342

PoiRank 0.297±0.331 0.314±0.365 0.241±0.361 0.372±0.428 0.309±0.340

Markov 0.255±0.323 0.237±0.323 0.335±0.430 0.291±0.405 0.200±0.287

Tmf 0.170±0.283 0.311±0.366 0.223±0.361 0.327±0.411 0.199±0.293

Gtmf 0.262±0.332 0.367±0.391 0.327±0.431 0.346±0.420 0.317±0.345

TGT mf 0.280±0.341 0.383±0.395 0.416±0.454 0.384±0.430 0.316±0.355

Rank+ Markov 0.299±0.335 0.333±0.373 0.388±0.438 0.365±0.423 0.281±0.332

Rank+ Tmf 0.289±0.333 0.323±0.374 0.287±0.397 0.376±0.428 0.270±0.332

Rank+ Gtmf 0.326±0.351 0.368±0.391 0.368±0.439 0.385±0.423 0.332±0.350

Rank+ TGT mf 0.337±0.355 0.383±0.395 0.452±0.458 0.402±0.432 0.356±0.364

start and end POIs. It is defined as follows:

$$TF1 = 2 \times TP \times TR$$

$$TP + TR(20)$$

- Tour pairs-F lscore: pairs-F1. The tour F lscore only considers individual POIs in the travel route, while the visiting order of a pair of POIs is ignored. We apply the tour pairs-F lscore, which was introduced in [5], to evaluate both POI identity and visiting order at the same time. The start and end POIs are included because they contain pairwise information. It is defined as:

$$l = 2 \times PP_{\text{Pair}} \times RP_{\text{Pair}}$$

$$PP_{\text{Pair}} + RP_{\text{Pair}}, (21)$$

where PP_{Pair} and RP_{Pair} are the precision and recall, respectively, of ordered POI pairs compared to the ground truth. Tour pairs-F l takes a value between 0 and 1 and will achieve a score of 1 if and only if the POIs and the visiting order are exactly the same as the user's real travel route.

Evaluation of Transition Weight Inference. To evaluate the performance of our proposed transition weight inference approaches, we normalize and randomly select 30% of the POIs from the observed matrix as the validation set, and then use different approaches to infer the unobserved transition weights (that is, new POIs are added to the dataset and no transition data can be observed). We adopt the following metric to evaluate the performance of the transition weight inference:

- Root-Mean-Square Error (RMSE) of Transition Weight: RMSE is a frequently used metric to measure the difference between a predicted value and the value actually observed. Let T in represent the inferred transition weight and T_{true} be the transition weight of

the real data. For each entry value u , taken from the set of all inferred entries U , we compute the RMSE of transition weight inference as:

$$\text{RMSE} = \frac{1}{|U|} \sqrt{\sum_{u \in U} (T_{in} - T_{tru})^2} \quad (22)$$

Questionnaire for Real Tourists. For each user query, we recommend two travel routes, generated by two different approaches. One is to recommend travel routes based only on POI popularity, which can be regarded as the cold start scenario of **Personalized**. The other is our proposed combination method, in which we infer the transition reward by employing all tourists' travel route data and apply them on the POI popularity baseline. We set the trade-off parameter $\alpha=0.7$; the reason for this setting will be explained below. Besides the recommended travel routes, a questionnaire was provided and we ask for feedback of each recommendation result. It requires users to give a rating score, ranging from 1 to 10, for each recommended route and select one with more reference value for their tour.

5.3 Results

For the quantitative evaluation results, the travel route recommendation performance of various approaches with a route length constraint is summarized in Tables 2 and 3. Each table is divided into three parts: representing location-based, transition-based, and combination methods. Performance with a travel time constraint is summarized in Tables 4 and 5. Performance on the two types of constraints is evaluated using the tour F1-score and the tour pairs-F1-score measurements. The best method for each dataset (city) is shown in bold; the second best is shown in italic. The performance of transition weight inference in terms of RMSE is shown in Table 6; the best result method for each dataset is shown in bold. Finally, for real user evaluation, the ratingscores are summarized in Fig. 9.

Comparison under the Route Length Budget. Under the route length constraint, as shown in Tables 2 and 3, our proposed combination method **Rank+TGT** outperforms all other baseline methods on different datasets in terms of both the tour F1-score and tour pairs-F1-scores. This result verifies that our proposed framework is efficient by considering both location and transition rewards, and transition knowledge is helpful to improve the performance of the travel route recommendation.

Although **PoiRank** uses machine learning method **RankSVM** with more features, it is not always better than baseline **PoiPop**, which indicates that POI popularity is more efficient than other features. For instance, most tourists will not miss visiting famous landmarks of the target travel city.

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2481

Table 3 Performance comparison of travel route recommendation under the route length budget in terms of tour pairs-F1-score (pairs-F1).

Budapest Edinburgh Glasgow Toronto Vienna
PoiPop 0.315±0.169 0.436±0.257 0.506±0.296 0.386±0.201 0.400±0.222

PoiRank 0.372 ± 0.226 0.424 ± 0.248 0.432 ± 0.241 0.508 ± 0.295 0.396 ± 0.209
 Markov 0.348 ± 0.235 0.369 ± 0.215 0.501 ± 0.300 0.454 ± 0.277 0.321 ± 0.181
 Tmf 0.292 ± 0.204 0.423 ± 0.246 0.418 ± 0.253 0.480 ± 0.281 0.320 ± 0.205
 Gtmf 0.357 ± 0.247 0.465 ± 0.275 0.494 ± 0.304 0.494 ± 0.282 0.403 ± 0.228
 TGT mf 0.367 ± 0.253 0.475 ± 0.282 0.555 ± 0.322 0.521 ± 0.294 0.407 ± 0.246
 Rank+ Markov 0.379 ± 0.231 0.437 ± 0.259 0.534 ± 0.301 0.502 ± 0.289 0.380 ± 0.206
 Rank+ Tmf 0.370 ± 0.253 0.429 ± 0.254 0.464 ± 0.280 0.514 ± 0.292 0.370 ± 0.208
 Rank+ Gtmf 0.401 ± 0.256 0.466 ± 0.276 0.521 ± 0.309 0.520 ± 0.290 0.415 ± 0.228
 Rank+ TGT mf 0.411 ± 0.262 0.475 ± 0.280 0.580 ± 0.327 0.533 ± 0.297 0.435 ± 0.247

Table 4 Performance comparison of travel route recommendation under the travel time budget in terms of tour F1score (TF1).

Budapest Edinburgh Glasgow Toronto Vienna

PersTour 0.372 ± 0.389 0.358 ± 0.391 0.539 ± 0.458 0.410 ± 0.438 0.320 ± 0.360
 PersTour+ Tmf 0.343 ± 0.405 0.378 ± 0.425 0.532 ± 0.459 0.425 ± 0.444 0.365 ± 0.374
 PersTour+ GTmf 0.394 ± 0.401 0.430 ± 0.401 0.541 ± 0.458 0.450 ± 0.449 0.379 ± 0.377
 PersTour+ TGT mf 0.423 ± 0.402 0.467 ± 0.412 0.561 ± 0.458 0.454 ± 0.449 0.395 ± 0.385

Table 5 Performance comparison of travel route recommendation under the travel time budget in terms of tour pairs-F1score (pairs-F1).

Budapest Edinburgh Glasgow Toronto Vienna

PersTour 0.387 ± 0.342 0.412 ± 0.321 0.596 ± 0.383 0.511 ± 0.336 0.325 ± 0.309
 PersTour+ Tmf 0.356 ± 0.321 0.442 ± 0.308 0.607 ± 0.366 0.513 ± 0.342 0.382 ± 0.323
 PersTour+ GTmf 0.416 ± 0.329 0.471 ± 0.331 0.616 ± 0.365 0.538 ± 0.345 0.386 ± 0.312
 PersTour+ TGT mf 0.434 ± 0.352 0.504 ± 0.343 0.627 ± 0.373 0.540 ± 0.348 0.402 ± 0.326

Among all of the transition-based approaches, Tmf, directly factorizing the weighted transition matrix, had no effect on the explicit feature pair factorization method Markov, which indicates that the latent location features might include those evident feature pairs. TGTmf achieves the best result by using global extra spatial and temporal features. Specifically, we observe a significant improvement in performance over Tmf when incrementally introducing additional spatial and temporal influences.

Comparison under the Travel Time Budget. Performance under the travel time constraint is shown in Tables 4 and 5. PersTour+ TGTmf, which combines spatial and temporal feature enhanced transition knowledge with personalized interest, achieves the best result. Again, we observe that performance is improved when additional spatial and temporal influences are incrementally introduced. In particular, the performance of Tmf noticeably improved when spatial influence features were introduced. This indicates that the location-location relation relies heavily on the distance between locations in the tour planning task.

Both our proposed PersTour+ GTmf and PersTour+ TGTmf consistently outperform PersTour, which only uses time-based user interest location rewards, in terms of the F1 and pairs-F1 scores. This verifies our idea of introducing transition knowledge to recommend travel routes.

Comparison with Transition Weight Inference. We use the RMSE of transition weights to evaluate different transition weight inference approaches, applied to all five

Table 6 Performance comparison of transition weight inference in terms of RMSE.

Budapest Edinburgh Glasgow Toronto Vienna
 Markov .10234 .09195 .10486 .08776 .09423
 Tmf .09874 .08995 .09064 .10056 .10285
 GTmf .08123 .08003 .07301 .08270 .08456
 TGT mf .07923 .07906 .07208 .08105 .08023

datasets; the results are shown in Table 6. Similar to the results on the performance of the travel route recommendation introduced above, the naive direct factorization method Tmf is not efficient enough when compared with the explicit feature pair factorization method Markov. However, the performance notably improves when the additional spatial and temporal features are introduced. In particular, spatial influence is the most effective factor in the factorization and transition weight inference, which is consistent with the results on the performance of recommendation.

For reference, Fig. 8 shows a visualization of transition weight inference from applying our proposed method to the tourism dataset of Kyoto. Each red point represents a POI and each black edge represents a transition. The thicker a black edge, the more users travel between the two POIs. The visualization of the observed transitions is shown in Fig. 8 (a), from which we can easily find the location of several landmarks because they have more links to other POIs. For instance, Fushimi Inari shrine, located in the lower right corner, is a famous sightseeing attraction in Kyoto. Figure 8 (b) shows the inferred transition weight graph, which

2482 IEICE TRANS. INF. & SYST., VOL. E102-D, NO. 12 DECEMBER 2019

Fig. 7 Impact of the trade-off parameter α on different datasets.

Fig. 8 The visualization of observed transition weights and inferred transition weights in the tourism dataset of Kyoto.

fills unobserved transition weights according to the observed transitions.

Impact of the Trade-off Parameter. As introduced in Sect. 4.3, we combine location and transition rewards by using a trade-off parameter α . This parameter controls the relative weights of location and location reward. The larger the value of α , the more the tour planning depends on locations, as opposed to transitions. Because this is a very important parameter that directly affects the final travel route planning procedure, we have to discover the effects of different values. Therefore, in order to understand the importance of the reward for each part, we evaluate the impact of α by plotting recommendation performance under different α values.

Figure 7 plots the impact of the trade-off parameter in terms of the recommendation performance F

1 score under

the route length budget. We take the performance produced by PoiRank as the baseline that only uses location reward, which will not change with the value of α . Then various types of transition knowledge produced by Markov, Tmf, GTmf, and TGT mf are included upon the PoiRank baseline.

According to the plot results, we observe that the importance of location and transition varied in different datasets (cities). This variation might demonstrate the different travel styles of different cities. For instance, transition patterns are more emphasized in Edinburgh and Glasgow, according to Figs. 7 (b) and 7 (c), because the performance decreases

as α increases. This can also be verified from Fig. 2, which shows that POIs in central Edinburgh are tightly linked to each other. In contrast, cities such as Budapest and Toronto rely more on location rewards, according to Figs. 7 (a) and 7 (d); the effects of transition reward on these datasets are not obvious.

Another interesting finding from these figures is that the trends of the different curves by introducing transition patterns are consistent with each other. The only difference might be caused by the inference ability of different methods. For instance, all F1 scores in Figs. 7 (a) and 7 (d) gradually increase as α rises, while the method Rank+TGT mf achieves the best result. The reason might be the higher transition weight inference ability of our proposed method TGT mf, which includes additional spatial and temporal features.

Finally, among all different values of α on different datasets, we find that a value between 0.5 and 0.7 can always achieve a higher performance upon the location reward. This could be an important hint for developing real travel route recommendation systems with our proposed framework.

Evaluation of Real Tourists We collected feedback from seven foreign tourists according to the experimental setting described in Sect. 5.2 and calculated simple statistics for the results. The average rating scores of two examined

SUN et al.: USER TRANSITION PATTERN ANALYSIS FOR TRAVEL ROUTE RECOMMENDATION

2483

Fig. 9 User evaluation of the recommended travel routes.

Fig. 10 Two recommended travel routes for query q1.

queries are summarized in Fig. 9 (a) and the users' selection is shown in Fig. 9 (b). As shown in Fig. 9, the average rating scores of the routes recommended by adding transition knowledge are higher than those based only on POI popularity, on both queries. Therefore, almost all users selected the routes recommended by our proposed combination method, which were considered more helpful for their trip. To better illustrate the effectiveness of our approach, we use a case study with query q

1 to see how transi-

tion knowledge affects the recommendation result compared with the location-based baseline method, which is presented in Fig. 10. The query q

1 implies that a user is departing from

Kyoto National Museum to the hotel near Sanjo, hoping for

a four-hour tour. Both routes are plotted on the Google map: red points indicate the most popular POIs and purple points represent relatively unpopular POIs. As shown by the plot-

ted results, the main differences are caused by the visiting orders. Without transition knowledge, it maximized POI rewards and suggested users to visit Heian Shrine after visit-

ing Maruyama Park. However, with transition knowledge,

it suggested users to travel through Maruyama Park, Yasaka

Shrine, and Gion in a straight line, which looks more natural

and suitable for tourists. Also, our method changed the visiting order and suggested users to travel through Ninenzaka

and Sanneizaka to Kiyomizu-dera, which is correct in reality because tourists have to climb these paths before entering Kiyomizu-dera.

6. Conclusion

This paper has proposed a technique that combines locations and transitions extracted from travel route data to recommend sightseeing routes according to user queries. To infer the city's transition patterns, which represent location–location relations, we enhance the latent factorization model

of a weighted transition matrix with additional spatial and temporal features, to enrich the latent features of location–location relations.

Real tourist travel route datasets were adopted for both qualitative and quantitative evaluation. Two types of travel budget constraints were considered to examine the generality of our proposed method. Comparing our method with recently published work, our method outperforms others in both qualitative and quantitative evaluations. Through the analysis on the recommendation performance and trade-off parameter, we reveal the efficiency by introducing transition patterns into the travel route recommendation task.

Acknowledgments

This work is partly supported by MIC SCOPE (172307001) and JSPS KAKENHI (16K12532).

References

- [1] W. Souffriau and P. Vansteenwegen, "Tourist trip planning functionalities: State-of-the-art and future," *International Conference on Web Engineering*, vol. 6385, pp. 474–485, Springer, 2010.
- [2] P. Vansteenwegen, W. Souffriau, and D. Van Oudheusden, "The orienteering problem: A survey," *European Journal of Operational Research*, vol. 209, no. 1, pp. 1–10, 2011.
- [3] C. Zhuang, N. J. Yuan, R. Song, X. Xie, and Q. Ma, "Understanding people lifestyles: Construction of urban movement knowledge graph from gps trajectory," *Proc. Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3616–3623, 2017.
- [4] C. Wang, Y. Gao, X. Gao, B. Yao, and G. Chen, "etour: A two-layer framework for tour recommendation with super-pois," *International Conference on Service-Oriented Computing*, vol. 11236, pp. 771–778, Springer, 2018.
- [5] D. Chen, C. S. Ong, and L. Xie, "Learning points and routes to recommend trajectories," *Proc. 25th ACM International Conference on Information and Knowledge Management*, pp. 2227–2232, ACM, 2016.
- [6] D. Gavalas, C. Konstantopoulos, K. Mastakas, G. Pantziou, and N. Vathis, "Efficient metaheuristics for the mixed team orienteering problem with time windows," *Algorithms*, vol. 9, no. 1, p. 6, 2016.
- [7] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "A survey on algorithmic approaches for solving tourist trip design problems," *Journal of Heuristics*, vol. 20, no. 3, pp. 291–328, 2014.
- [8] Z. Friggstad, S. Gollapudi, K. Kollias, T. Sarlos, C. Swamy, and A. Tomkins, "Orienteering algorithms for generating travel itineraries," *Proc. Eleventh ACM International Conference on Web Search and Data Mining*, pp. 180–188, ACM, 2018.
- [9] P. Vansteenwegen, W. Souffriau, G. V. Berghe, and D. Van Oudheusden, "Iterated local search for the team orienteering problem with time windows," *Computers & Operations Research*, vol. 36, no. 12, pp. 3281–3290, 2009.

- [10] A. Garcia, P. Vansteenwegen, O. Arbelaitz, W. Souffriau, and M.T. 2484 IEICE TRANS. INF. & SYST., VOL.E102-D, NO.12 DECEMBER 2019
Linaza, "Integrating public transportation in personalised electronic tourist guides," *Computers & Operations Research*, vol.40, no.3, pp.758-774, 2013.
- [11] D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, N. Vathis, and C. Zaroliagis, "The ecompass multimodal tourist tour planner," *Expert systems with Applications*, vol.42, no.21, pp.7303-7316, 2015.
- [12] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu, "Automatic construction of travel itineraries using social breadcrumbs," *Proc. 21st ACM conference on Hypertext and hypermedia*, pp.35-44, ACM, 2010.
- [13] K.H. Lim, J. Chan, C. Leckie, and S. Karunasekera, "Personalized tour recommendation based on user interests and points of interest visit durations," *Proc. Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp.1778-1784, 2015.
- [14] I.R. Brilhante, J.A. Macedo, F.M. Nardini, R. Perego, and C. Renso, "On planning sightseeing tours with tripbuilder," *Information Processing & Management*, vol.51, no.2, pp.1-15, 2015.
- [15] P. Zhao, X. Xu, Y. Liu, V.S. Sheng, K. Zheng, and H. Xiong, "Photo2trip: Exploiting visual contents in geo-tagged photos for personalized tour recommendation," *Proc. 2017 ACM on Multimedia Conference*, pp.916-924, ACM, 2017.
- [16] P. Vansteenwegen, W. Souffriau, G.V. Berghe, and D. Van Oudheusden, "The city trip planner: an expert system for tourists," *Expert Systems with Applications*, vol.38, no.6, pp.6540-6546, 2011.
- [17] Y. Kurata, "Ct-planner2: More flexible and interactive assistance for day tour planning," *ENTER*, pp.25-37, 2011.
- [18] C. Zhang, H. Liang, and K. Wang, "Trip recommendation meets real-world constraints: Poi availability, diversity, and traveling time uncertainty," *ACM Transactions on Information Systems (TOIS)*, vol.35, no.1, p.5, 2016.
- [19] Y.-T. Wen, J. Yeo, W.-C. Peng, and S.-W. Hwang, "Efficient keyword-aware representative travel route recommendation," *IEEE Trans. Knowl. Data Eng.*, vol.29, no.8, pp.1639-1652, 2017.
- [20] J. Sun, C. Zhuang, and Q. Ma, "Travel route recommendation by considering user transition patterns," *e-Review of Tourism Research*, vol.16, no.2/3, 2019.
- [21] C.H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*, 1998.
- [22] C.-P. Lee and C.-J. Lin, "Large-scale linear ranksvm," *Neural computation*, vol.26, no.4, pp.781-817, 2014.
- [23] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol.42, no.8, pp.30-37, 2009.
- [24] 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," *Proc. 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.831-840, ACM, 2014.
- [25] M. Nickel, V. Tresp, and H.-P. Kriegel, "Factorizing yago: scalable machine learning for linked data," *Proc. 21st international conference on World Wide Web*, pp.271-280, ACM, 2012.
- [26] D. Yang, D. Zhang, Z. Yu, and Z. Wang, "A sentiment-enhanced personalized location recommendation system," *Proc. 24th ACM Conference on Hypertext and Social Media*, pp.119-128, ACM, 2013.

[27] R.W. Sinnott, "Virtues of the haversine," Sky Telesc., vol.68, p.159, 1984.

[28] X. Liu, Y. Liu, K. Aberer, and C. Miao, "Personalized point-of-interest recommendation by mining users' preference transition," Proc. 22nd ACM international conference on Information & Knowledge Management, pp.733–738, ACM, 2013.

[29] C. Verbeeck, P. Vansteenwegen, and E.H. Aghezzaf, "An extension of the arc orienteering problem and its application to cycle trip planning," Transportation research part E: logistics and transportation review, vol.68, pp.64–78, 2014.

[30] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," Appears in the International Joint Conference on Artificial Intelligence (IJCAI), 1995, pp.1137–1145, Montreal, Canada, 1995.

[31] Z. Yao, Y. Fu, B. Liu, Y. Liu, and H. Xiong, "Poi recommendation: A temporal matching between poi popularity and user regularity," IEEE 16th International Conference on Data Mining, pp.549–558, 2016.

Junjie Sun received his master degree and became a Ph.D. candidate in Department of Social Informatics, Graduate School of Informatics, Kyoto University in 2019. His current research interests include knowledge discovery, multimedia mining, recommender system, and urban computing.

Chenyi Zhuang joined Artificial Intelligence Research Center (AIRC), AIST as a researcher in April 2018. Prior to that, during October 2017 and March 2018, he was a post-doctor in Kyoto University. He received the B.S. degree in SE from Nanjing University in 2011, the M.S. degree and Ph.D. degree in Informatics from Kyoto University in 2014 and 2017, respectively. In between, from 2015 to 2018, he was also serving as a young scientist in Japan Society for the Promotion of Science (JSPS). His current research primarily involves structured data mining, machine learning, and urban computing.

Qiang Ma received his Ph.D. degree from Department of Social Informatics, Graduate School of Informatics, Kyoto University in 2004. He was a research fellow (DC2) of JSPS from 2003 to 2004. He joined National Institute of Information and Communications Technology as a research fellow in 2004. From 2006 to 2007, he served as an assistant manager at NEC. From October 2007, he joined Kyoto University and has been an associate professor since August 2010. His general research interests are in the area of databases and information retrieval. His current interests include Web information system, Web mining, knowledge discovery, and multimedia mining and search.

View publication stats

In []: `#all images extract`

```
In [10]: for i in page.images:
         with open(i.name, 'wb') as f:
             f.write(i.data)
```

```
In [8]: from PyPDF2 import PdfReader
reader= PyPDF2.PdfReader('Final nm.pdf')
metadata= reader.metadata
print("Metadata", metadata)
```

```
Metadata {'/Creator': 'pdftk 1.44 - www.pdftk.com', '/Producer': 'itext-paul
o-155 (itextpdf.sf.net-lowagie.com)', '/ModDate': "D:20191203151321+09'00'",
'/CreationDate': "D:20191203151321+09'00'", '/rgid': 'PB:335060973_AS:832284
164882432@1575443416111'}
```

2nd method

```
In [5]: #pip install pdfplumber
```

```
In [11]: import pdfplumber
```

```
In [13]: with pdfplumber.open('Final nm.pdf') as f:
         for i in f.pages:
             print(i.extract_tables())
```



```

-----
ValueError                                Traceback (most recent call last)
Cell In[18], line 10
      7         df= pd.DataFrame(table[1:], columns= table[0]) #first row =
header
      9 #combined all tables into a dataframe
--> 10 final_df= pd.concat(tables_list, ignore_index= True)
      11 final_df

File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\reshape\concat.py:382, in concat(objs, axis, join, ignore_index, keys, levels, names, verify_integrity, sort, copy)
    379 elif copy and using_copy_on_write():
    380     copy = False
--> 382 op = _Concatenator(
    383     objs,
    384     axis=axis,
    385     ignore_index=ignore_index,
    386     join=join,
    387     keys=keys,
    388     levels=levels,
    389     names=names,
    390     verify_integrity=verify_integrity,
    391     copy=copy,
    392     sort=sort,
    393 )
    395 return op.get_result()

File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\reshape\concat.py:445, in _Concatenator.__init__(self, objs, axis, join, keys, levels, names, ignore_index, verify_integrity, copy, sort)
    442 self.verify_integrity = verify_integrity
    443 self.copy = copy
--> 445 objs, keys = self._clean_keys_and_objs(objs, keys)
    447 # figure out what our result ndim is going to be
    448 ndims = self._get_ndims(objs)

File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\reshape\concat.py:507, in _Concatenator._clean_keys_and_objs(self, objs, keys)
    504     objs_list = list(objs)
    506 if len(objs_list) == 0:
--> 507     raise ValueError("No objects to concatenate")
    509 if keys is None:
    510     objs_list = list(com.not_none(*objs_list))

ValueError: No objects to concatenate

```

```

In [19]: with pdfplumber.open("Final nm.pdf") as pdf:
        for page_num, page in enumerate(pdf.pages, start=1):
            tables = page.extract_tables()
            if tables:
                for t in tables:
                    df = pd.DataFrame(t[1:], columns=t[0])
                    print(f"Table from Page {page_num}:")
                    print(df.head())

```

```
else:  
    print(f"No tables found on page {page_num}")
```

Table from Page 1:

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/335060973>
User Transition Pattern Analysis for Travel Route Recommendation
Article in IEICE Transactions on Information and Systems · August 2019
DOI: 10.1587/transinf.2019EDP7096
CITATIONS 325
3 authors: Junjie Sun, Chenyi Zhuang, Kyoto University
PUBLICATIONS 22 CITATIONS 21 PUBLICATIONS 598 CITATIONS
SEE PROFILE SEE PROFILE
Qiang Ma, Kyoto Institute of Technology
130 PUBLICATIONS 1,053 CITATIONS
SEE PROFILE

All content following this page was uploaded by...

None

0

No tables found on page 2

No tables found on page 3

No tables found on page 4

No tables found on page 5

No tables found on page 6

Table from Page 7:

None

0 None

Table from Page 8:

Empty DataFrame

Columns: [],]

Index: []

Table from Page 9:

None

0

None

1

None

2

None

None

None

None

None

3

None

None

None

4

None

None

Table from Page 9:

None None

0

None

1

None

None

2

None

None

3

None

4

None

None

No tables found on page 10

No tables found on page 11

No tables found on page 12

Table from Page 13:

Empty DataFrame

Columns: [],]

Index: []

Table from Page 13:

Empty DataFrame

Columns: [],]

Index: []

Table from Page 13:

0

Table from Page 13:

0

None

None

No tables found on page 14

No tables found on page 15

```
In [20]: with pdfplumber.open("Final nm.pdf") as pdf:
          page = pdf.pages[0]
          words = page.extract_words()
          print(words)    # shows text + coordinates
```

[{'text': 'See', 'x0': 39.75500858354123, 'x1': 48.84838448553806, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 9.09337590199683, 'direction': 'ltr'}, {'text': 'discussions,', 'x0': 50.05093469430181, 'x1': 80.21660038708653, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 30.16566569278472, 'direction': 'ltr'}, {'text': 'stats,', 'x0': 81.41915059585025, 'x1': 94.87140658370366, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 13.452255987853405, 'direction': 'ltr'}, {'text': 'and', 'x0': 96.0739567924674, 'x1': 105.6943582836798, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 9.620401491212405, 'direction': 'ltr'}, {'text': 'author', 'x0': 106.89690849244356, 'x1': 123.75008159463799, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 16.853173102194432, 'direction': 'ltr'}, {'text': 'profiles', 'x0': 124.95263180340174, 'x1': 143.68096775903933, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 18.72833595563759, 'direction': 'ltr'}, {'text': 'for', 'x0': 144.8835179678031, 'x1': 151.8658793061646, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 6.9823613383615, 'direction': 'ltr'}, {'text': 'this', 'x0': 153.06842951492834, 'x1': 162.28992211398946, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 9.221492599061122, 'direction': 'ltr'}, {'text': 'publication', 'x0': 163.49247232275323, 'x1': 192.01882654099515, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 28.52635421824192, 'direction': 'ltr'}, {'text': 'at:', 'x0': 193.2213767497589, 'x1': 199.8543287468558, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 6.63295199709691, 'direction': 'ltr'}, {'text': 'https://www.researchgate.net/publication/335060973', 'x0': 201.05267361780488, 'x1': 335.1501225948452, 'top': 73.8476460565596, 'doctop': 73.8476460565596, 'bottom': 79.81089734409079, 'upright': True, 'height': 5.963251287531193, 'width': 134.0974489770403, 'direction': 'ltr'}, {'text': 'User', 'x0': 39.75500858354123, 'x1': 70.12126098811397, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 30.36625240457274, 'direction': 'ltr'}, {'text': 'Transition', 'x0': 73.18182722480704, 'x1': 141.62721794634896, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 68.44539072154193, 'direction': 'ltr'}, {'text': 'Pattern', 'x0': 144.68778418304203, 'x1': 194.11625258428904, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 49.42846840124702, 'direction': 'ltr'}, {'text': 'Analysis', 'x0': 197.1768188209821, 'x1': 253.27641592756544, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 56.09959710658333, 'direction': 'ltr'}, {'text': 'for', 'x0': 256.33698216425853, 'x1': 275.7097841654918, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 19.372802001233254, 'direction': 'ltr'}, {'text': 'Travel', 'x0': 278.7703504021848, 'x1': 319.3794683117376, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 40.609117909552765,

'direction': 'ltr'}, {'text': 'Route', 'x0': 322.4400345484306, 'x1': 360.97211082184407, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 38.53207627341345, 'direction': 'ltr'}, {'text': 'Recommendation', 'x0': 364.03267705853716, 'x1': 481.34359929653186, 'top': 98.87764280116471, 'doctop': 98.87764280116471, 'bottom': 112.12931232901178, 'upright': True, 'height': 13.25166952784707, 'width': 117.3109222379947, 'direction': 'ltr'}, {'text': 'Article', 'x0': 39.75500858354123, 'x1': 58.25752269298643, 'top': 128.35937217176763, 'doctop': 128.35937217176763, 'bottom': 134.9852069356917, 'upright': True, 'height': 6.625834763923535, 'width': 18.502514109445194, 'direction': 'ltr'}, {'text': 'in', 'x0': 60.92662122764067, 'x1': 65.97364380172306, 'top': 128.49525359845586, 'doctop': 128.49525359845586, 'bottom': 135.1210883623794, 'upright': True, 'height': 6.625834763923535, 'width': 5.0470225740823835, 'direction': 'ltr'}, {'text': 'IEICE', 'x0': 68.63950700752042, 'x1': 82.88440467093298, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 14.244897663412559, 'direction': 'ltr'}, {'text': 'Transactions', 'x0': 84.22057156955937, 'x1': 119.6532583225807, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 35.43268675302133, 'direction': 'ltr'}, {'text': 'on', 'x0': 120.98942522120709, 'x1': 128.20407930200483, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 7.214654080797743, 'direction': 'ltr'}, {'text': 'Information', 'x0': 129.5402462006312, 'x1': 162.62412627761117, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 33.083880076979966, 'direction': 'ltr'}, {'text': 'and', 'x0': 163.96029317623754, 'x1': 174.64962816647355, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 10.689334990236006, 'direction': 'ltr'}, {'text': 'Systems', 'x0': 175.98579506509992, 'x1': 199.11474615313512, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 23.128951088035194, 'direction': 'ltr'}, {'text': '.', 'x0': 200.44250155452872, 'x1': 202.08925445771402, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 1.6467529031853019, 'direction': 'ltr'}, {'text': 'August', 'x0': 203.42542135634042, 'x1': 222.552342737502, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 19.126921381161566, 'direction': 'ltr'}, {'text': '2019', 'x0': 223.88850963612836, 'x1': 237.04959154618325, 'top': 128.43378341030882, 'doctop': 128.43378341030882, 'bottom': 135.05961817423236, 'upright': True, 'height': 6.625834763923535, 'width': 13.16108191005489, 'direction': 'ltr'}, {'text': 'DOI:', 'x0': 39.75500858354123, 'x1': 48.05509602249258, 'top': 139.59740911664267, 'doctop': 139.59740911664267, 'bottom': 144.23549345138917, 'upright': True, 'height': 4.638084334746509, 'width': 8.300087438951344, 'direction': 'ltr'}, {'text': '10.1587/transinf.2019EDP7096', 'x0': 48.99041285153104, 'x1': 107.8791367285482, 'top': 139.59740911664267, 'doctop': 139.59740911664267, 'bottom': 144.23549345138917, 'upright': True, 'height': 4.638084334746509, 'width': 58.88872387701716, 'direction': 'ltr'}, {'text': 'CITATIONS', 'x0': 39.75500858354123, 'x1': 63.47860292493643, 'top': 171.6557082730518, 'doctop': 171.6557082730518, 'bottom': 176.95637608419065, 'upright': True, 'height': 5.30066781138851, 'width': 23.7235943413952, 'direction': 'ltr'}, {'text': 'READS', 'x0': 297.4999809001669, 'x1': 312.34081549792546, 'top': 171.6557082730518, 'doctop': 171.6557082730518, 'bottom': 176.95637608419065, 'upright': True,

'height': 5.300667811138851, 'width': 14.840834597758544, 'direction': 'ltr'}, {'text': '7', 'x0': 39.75500858354123, 'x1': 43.7033331565577, 'top': 180.6238791480647, 'doctop': 180.6238791480647, 'bottom': 188.57488086477292, 'upright': True, 'height': 7.95100171670822, 'width': 3.948324573016464, 'direction': 'ltr'}, {'text': '325', 'x0': 297.4999809001669, 'x1': 309.3449546192163, 'top': 180.6238791480647, 'doctop': 180.6238791480647, 'bottom': 188.57488086477292, 'upright': True, 'height': 7.95100171670822, 'width': 11.844973719049392, 'direction': 'ltr'}, {'text': '3', 'x0': 39.75500858354123, 'x1': 43.15204299183979, 'top': 206.54422238606537, 'doctop': 206.54422238606537, 'bottom': 213.1700571499889, 'upright': True, 'height': 6.625834763923535, 'width': 3.397034408298559, 'direction': 'ltr'}, {'text': 'authors:', 'x0': 44.507621466057344, 'x1': 68.4550967743709, 'top': 206.54422238606537, 'doctop': 206.54422238606537, 'bottom': 213.1700571499889, 'upright': True, 'height': 6.625834763923535, 'width': 23.947475308313557, 'direction': 'ltr'}, {'text': 'Junjie', 'x0': 66.2583476392354, 'x1': 83.21116705474289, 'top': 224.50838748720014, 'doctop': 224.50838748720014, 'bottom': 231.13422225112367, 'upright': True, 'height': 6.625834763923535, 'width': 16.95281941550749, 'direction': 'ltr'}, {'text': 'Sun', 'x0': 84.54733395336928, 'x1': 95.3110801821465, 'top': 224.50838748720014, 'doctop': 224.50838748720014, 'bottom': 231.13422225112367, 'upright': True, 'height': 6.625834763923535, 'width': 10.763746228777222, 'direction': 'ltr'}, {'text': 'Chenyi', 'x0': 324.00331995586106, 'x1': 343.0170069472991, 'top': 224.50838748720014, 'doctop': 224.50838748720014, 'bottom': 231.13422225112367, 'upright': True, 'height': 6.625834763923535, 'width': 19.01368699143802, 'direction': 'ltr'}, {'text': 'Zhuang', 'x0': 344.3531738459254, 'x1': 365.4827279437273, 'top': 224.50838748720014, 'doctop': 224.50838748720014, 'bottom': 231.13422225112367, 'upright': True, 'height': 6.625834763923535, 'width': 21.129554097801872, 'direction': 'ltr'}, {'text': 'Kyoto', 'x0': 66.2583476392354, 'x1': 82.38293777551078, 'top': 235.10972310947773, 'doctop': 235.10972310947773, 'bottom': 241.73555787340126, 'upright': True, 'height': 6.625834763923535, 'width': 16.124590136275387, 'direction': 'ltr'}, {'text': 'University', 'x0': 83.71910467413717, 'x1': 111.55861069471605, 'top': 235.10972310947773, 'doctop': 235.10972310947773, 'bottom': 241.73555787340126, 'upright': True, 'height': 6.625834763923535, 'width': 27.83950602057888, 'direction': 'ltr'}, {'text': 'Kyoto', 'x0': 324.00331995586106, 'x1': 340.1279100921365, 'top': 235.10972310947773, 'doctop': 235.10972310947773, 'bottom': 241.73555787340126, 'upright': True, 'height': 6.625834763923535, 'width': 16.124590136275458, 'direction': 'ltr'}, {'text': 'University', 'x0': 341.46407699076286, 'x1': 369.3035830113418, 'top': 235.10972310947773, 'doctop': 235.10972310947773, 'bottom': 241.73555787340126, 'upright': True, 'height': 6.625834763923535, 'width': 27.839506020578938, 'direction': 'ltr'}, {'text': '7', 'x0': 66.2583476392354, 'x1': 69.3156786067041, 'top': 247.37754675666918, 'doctop': 247.37754675666918, 'bottom': 253.34079804420037, 'upright': True, 'height': 5.963251287531193, 'width': 3.057330967468701, 'direction': 'ltr'}, {'text': 'PUBLICATIONS', 'x0': 70.64796317033473, 'x1': 103.90395946933813, 'top': 247.8528080581725, 'doctop': 247.8528080581725, 'bottom': 253.15347586931136, 'upright': True, 'height': 5.300667811138851, 'width': 33.2559962990034, 'direction': 'ltr'}, {'text': '22', 'x0': 107.90793085058601, 'x1': 114.02259278552341, 'top': 247.37754675666918, 'doctop': 247.37754675666918, 'bottom': 253.34079804420037, 'upright': True, 'height': 5.963251287531193, 'width': 6.114661934937402, 'direction': 'ltr'}, {'text': 'CITATIONS', 'x0': 115.35164209318137, 'x1': 139.07523643457657, 'top': 247.8528080581725, 'doctop': 247.8528080581725, 'bottom': 253.15347586931136, 'upright': True, 'height': 5.300667811138851, 'width': 23.7235943413952, 'direction': 'ltr'}, {'text': '21', 'x0': 324.00331995586106, 'x1': 330.1179818907985, 'top': 247.37754675666918, 'doctop': 247.37754675666918, 'bottom': 253.34079804420037, 'upright': True, 'height': 5.963251287531193, 'width': 6.114661934937402, 'direction': 'ltr'}

ight': True, 'height': 5.963251287531193, 'width': 6.1146619349374305, 'direction': 'ltr'}, {'text': 'PUBLICATIONS', 'x0': 331.4470345113738, 'x1': 364.70303081037724, 'top': 247.8528080581725, 'doctop': 247.8528080581725, 'bottom': 253.15347586931136, 'upright': True, 'height': 5.300667811138851, 'width': 33.25599629900347, 'direction': 'ltr'}, {'text': '598', 'x0': 368.7069988787077, 'x1': 377.87899178111377, 'top': 247.37754675666918, 'doctop': 247.37754675666918, 'bottom': 253.34079804420037, 'upright': True, 'height': 5.963251287531193, 'width': 9.171992902406089, 'direction': 'ltr'}, {'text': 'CITATIONS', 'x0': 379.2048058327991, 'x1': 402.9284001741943, 'top': 247.8528080581725, 'doctop': 247.8528080581725, 'bottom': 253.15347586931136, 'upright': True, 'height': 5.300667811138851, 'width': 23.7235943413952, 'direction': 'ltr'}, {'text': 'SEE', 'x0': 73.54676587955129, 'x1': 81.96105841952243, 'top': 263.75481149158907, 'doctop': 263.75481149158907, 'bottom': 269.0554793027279, 'upright': True, 'height': 5.300667811138851, 'width': 8.414292539971143, 'direction': 'ltr'}, {'text': 'PROFILE', 'x0': 83.02999193842354, 'x1': 102.05338467790213, 'top': 263.75481149158907, 'doctop': 263.75481149158907, 'bottom': 269.0554793027279, 'upright': True, 'height': 5.300667811138851, 'width': 19.023392739478595, 'direction': 'ltr'}, {'text': 'SEE', 'x0': 331.291738196177, 'x1': 339.70603073614814, 'top': 263.75481149158907, 'doctop': 263.75481149158907, 'bottom': 269.0554793027279, 'upright': True, 'height': 5.300667811138851, 'width': 8.414292539971143, 'direction': 'ltr'}, {'text': 'PROFILE', 'x0': 340.7749642550492, 'x1': 359.7983569945278, 'top': 263.75481149158907, 'doctop': 263.75481149158907, 'bottom': 269.0554793027279, 'upright': True, 'height': 5.300667811138851, 'width': 19.023392739478595, 'direction': 'ltr'}, {'text': 'Qiang', 'x0': 66.2583476392354, 'x1': 82.63528885452115, 'top': 287.45381774447367, 'doctop': 287.45381774447367, 'bottom': 294.0796525083972, 'upright': True, 'height': 6.625834763923535, 'width': 16.376941215285754, 'direction': 'ltr'}, {'text': 'Ma', 'x0': 83.97145575314754, 'x1': 92.17610270691222, 'top': 287.45381774447367, 'doctop': 287.45381774447367, 'bottom': 294.0796525083972, 'upright': True, 'height': 6.625834763923535, 'width': 8.20464695376468, 'direction': 'ltr'}, {'text': 'Kyoto', 'x0': 66.2583476392354, 'x1': 82.38293777551078, 'top': 298.05515336675137, 'doctop': 298.05515336675137, 'bottom': 304.6809881306749, 'upright': True, 'height': 6.625834763923535, 'width': 16.124590136275387, 'direction': 'ltr'}, {'text': 'Institute', 'x0': 83.71910467413717, 'x1': 106.98717271649262, 'top': 298.05515336675137, 'doctop': 298.05515336675137, 'bottom': 304.6809881306749, 'upright': True, 'height': 6.625834763923535, 'width': 23.268068042355452, 'direction': 'ltr'}, {'text': 'of', 'x0': 108.323339615119, 'x1': 113.84918231269222, 'top': 298.05515336675137, 'doctop': 298.05515336675137, 'bottom': 304.6809881306749, 'upright': True, 'height': 6.625834763923535, 'width': 5.525842697573225, 'direction': 'ltr'}, {'text': 'Technology', 'x0': 115.18534921131861, 'x1': 147.10453170494426, 'top': 298.05515336675137, 'doctop': 298.05515336675137, 'bottom': 304.6809881306749, 'upright': True, 'height': 6.625834763923535, 'width': 31.919182493625655, 'direction': 'ltr'}, {'text': '130', 'x0': 66.2583476392354, 'x1': 75.4303405416415, 'top': 310.3229770139428, 'doctop': 310.3229770139428, 'bottom': 316.286228301474, 'upright': True, 'height': 5.963251287531193, 'width': 9.171992902406103, 'direction': 'ltr'}, {'text': 'PUBLICATIONS', 'x0': 76.75615459332676, 'x1': 110.01215089233014, 'top': 310.79823831544616, 'doctop': 310.79823831544616, 'bottom': 316.098906126585, 'upright': True, 'height': 5.300667811138851, 'width': 33.255996299003385, 'direction': 'ltr'}, {'text': '1,053', 'x0': 114.0161255864954, 'x1': 127.88476118932054, 'top': 310.3229770139428, 'doctop': 310.3229770139428, 'bottom': 316.286228301474, 'upright': True, 'height': 5.963251287531193, 'width': 13.868635602825137, 'direction': 'ltr'}, {'text': 'CITATIONS', 'x0': 129.21413407624502, 'x1': 152.93772841764022, 'top': 310.79823831544616, 'doctop': 310.79823831544616, 'bottom': 316.0989061

26585, 'upright': True, 'height': 5.300667811138851, 'width': 23.72359434139
52, 'direction': 'ltr'}, {'text': 'SEE', 'x0': 73.54676587955129, 'x1': 81.9
6105841952243, 'top': 326.7002417488626, 'doctop': 326.7002417488626, 'botto
m': 332.0009095600015, 'upright': True, 'height': 5.300667811138908, 'wid
h': 8.414292539971143, 'direction': 'ltr'}, {'text': 'PROFILE', 'x0': 83.029
99193842354, 'x1': 102.05338467790213, 'top': 326.7002417488626, 'doctop': 3
26.7002417488626, 'bottom': 332.0009095600015, 'upright': True, 'height': 5.
300667811138908, 'width': 19.023392739478595, 'direction': 'ltr'}, {'text':
'All', 'x0': 43.06792596550301, 'x1': 50.04964047662321, 'top': 804.93151280
69021, 'doctop': 804.9315128069021, 'bottom': 811.5573475708256, 'upright':
True, 'height': 6.625834763923535, 'width': 6.981714511120202, 'direction':
'ltr'}, {'text': 'content', 'x0': 51.38580737524959, 'x1': 72.8000653768221
6, 'top': 804.9315128069021, 'doctop': 804.9315128069021, 'bottom': 811.5573
475708256, 'upright': True, 'height': 6.625834763923535, 'width': 21.4142580
0157257, 'direction': 'ltr'}, {'text': 'following', 'x0': 74.13623227544853,
'x1': 99.88575309167675, 'top': 804.9315128069021, 'doctop': 804.93151280690
21, 'bottom': 811.5573475708256, 'upright': True, 'height': 6.62583476392353
5, 'width': 25.749520816228213, 'direction': 'ltr'}, {'text': 'this', 'x0':
101.22191999030315, 'x1': 111.46802287814884, 'top': 804.9315128069021, 'doc
top': 804.9315128069021, 'bottom': 811.5573475708256, 'upright': True, 'heig
ht': 6.625834763923535, 'width': 10.246102887845694, 'direction': 'ltr'},
{'text': 'page', 'x0': 112.80418977677522, 'x1': 126.30497485600075, 'top':
804.9315128069021, 'doctop': 804.9315128069021, 'bottom': 811.5573475708256,
'upright': True, 'height': 6.625834763923535, 'width': 13.500785079225537,
'direction': 'ltr'}, {'text': 'was', 'x0': 127.64114175462713, 'x1': 138.498
7107975367, 'top': 804.9315128069021, 'doctop': 804.9315128069021, 'bottom':
811.5573475708256, 'upright': True, 'height': 6.625834763923535, 'width': 1
0.85756904290956, 'direction': 'ltr'}, {'text': 'uploaded', 'x0': 139.834877
6961631, 'x1': 166.32527583397984, 'top': 804.9315128069021, 'doctop': 804.9
315128069021, 'bottom': 811.5573475708256, 'upright': True, 'height': 6.6258
34763923535, 'width': 26.490398137816754, 'direction': 'ltr'}, {'text': 'b
y', 'x0': 167.66144273260622, 'x1': 174.4037470825935, 'top': 804.9315128069
021, 'doctop': 804.9315128069021, 'bottom': 811.5573475708256, 'upright': Tr
ue, 'height': 6.625834763923535, 'width': 6.742304349987279, 'direction': 'l
tr'}, {'text': 'Junjie', 'x0': 175.73991755917064, 'x1': 192.69273697467813,
'top': 804.9315128069021, 'doctop': 804.9315128069021, 'bottom': 811.5573475
708256, 'upright': True, 'height': 6.625834763923535, 'width': 16.9528194155
0749, 'direction': 'ltr'}, {'text': 'Sun', 'x0': 194.0289038733045, 'x1': 20
4.79265010208175, 'top': 804.9315128069021, 'doctop': 804.9315128069021, 'bo
ttom': 811.5573475708256, 'upright': True, 'height': 6.625834763923535, 'wid
th': 10.76374622877725, 'direction': 'ltr'}, {'text': 'on', 'x0': 206.126228
75087428, 'x1': 213.34088283167205, 'top': 804.9315128069021, 'doctop': 804.
9315128069021, 'bottom': 811.5573475708256, 'upright': True, 'height': 6.625
834763923535, 'width': 7.2146540807977715, 'direction': 'ltr'}, {'text': '0
4', 'x0': 214.6770497302984, 'x1': 221.25759068532585, 'top': 804.9315128069
021, 'doctop': 804.9315128069021, 'bottom': 811.5573475708256, 'upright': Tr
ue, 'height': 6.625834763923535, 'width': 6.580540955027459, 'direction': 'l
tr'}, {'text': 'December', 'x0': 222.59375758395223, 'x1': 250.8603192614121
2, 'top': 804.9315128069021, 'doctop': 804.9315128069021, 'bottom': 811.5573
475708256, 'upright': True, 'height': 6.625834763923535, 'width': 28.2665616
7745989, 'direction': 'ltr'}, {'text': '2019.', 'x0': 252.1964861600385, 'x
1': 267.0043209732787, 'top': 804.9315128069021, 'doctop': 804.931512806902
1, 'bottom': 811.5573475708256, 'upright': True, 'height': 6.62583476392353
5, 'width': 14.80783481324022, 'direction': 'ltr'}, {'text': 'The', 'x0': 4
3.06792596550301, 'x1': 51.417513041476866, 'top': 820.9875151375587, 'docto
p': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'heigh

```
t': 5.300667811138851, 'width': 8.349587075973858, 'direction': 'ltr'}, {'text': 'user', 'x0': 52.48644656037797, 'x1': 62.0550837108714, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 9.56863715049343, 'direction': 'ltr'}, {'text': 'has', 'x0': 63.12401722977251, 'x1': 70.9404316513352, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 7.816414421562705, 'direction': 'ltr'}, {'text': 'requested', 'x0': 72.00936517023632, 'x1': 94.37414570855708, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 22.36478053832076, 'direction': 'ltr'}, {'text': 'enhancement', 'x0': 95.4430792274582, 'x1': 126.10309448310423, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 30.66001525564603, 'direction': 'ltr'}, {'text': 'of', 'x0': 127.17202800200533, 'x1': 131.5927021600639, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 4.420674158058574, 'direction': 'ltr'}, {'text': 'the', 'x0': 132.661635678965, 'x1': 139.9629950226817, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 7.301359343716683, 'direction': 'ltr'}, {'text': 'downloaded', 'x0': 141.0319285415828, 'x1': 168.89666931889352, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 27.864740777310715, 'direction': 'ltr'}, {'text': 'file.', 'x0': 169.9656028377946, 'x1': 178.11072084048135, 'top': 820.9875151375587, 'doctop': 820.9875151375587, 'bottom': 826.2881829486976, 'upright': True, 'height': 5.300667811138851, 'width': 8.145118002686758, 'direction': 'ltr'}]}
```

3rd format

```
In [22]: #pip install PyMuPDF
```

```
In [23]: import fitz
```

```
In [26]: doc = fitz.open('Final nm.pdf')
print(doc.page_count)
```

15

```
In [32]: print(doc.metadata)
```

```
{'format': 'PDF 1.4', 'title': '', 'author': '', 'subject': '', 'keywords': '', 'creator': 'pdftk 1.44 - www.pdftk.com', 'producer': 'itext-paulo-155 (itextpdf.sf.net-lowagie.com)', 'creationDate': "D:20191203151321+09'00'", 'modDate': "D:20191203151321+09'00'", 'trapped': '', 'encryption': None}
```

```
In [41]: tables = []

with pdfplumber.open('Final nm.pdf') as pdf:
    for page in pdf.pages:
        tables_on_page = page.extract_tables({
            'vertical_strategy': 'text',
            'horizontal_strategy': 'text'
```

```
    })
    if tables_on_page:
        for table in tables_on_page:
            if table:
                tables.append({
                    'page': pdf.pages.index(page) + 1,
                    'data': table
                })

for table in tables:
    print("Page:", table['page'])
    print(pd.DataFrame(table['data']))
```

Page: 1

0 \
0 See discussions, s
1
2 User T
3
4 Article in IEICE
5
6 DOI: 10.1587/transinf.2
7
8 CITATIONS
9
10 7
11
12 3 authors:
13
14 Junjie
15
16 Kyoto
17
18 7 PUBL
19
20 SEE
21
22 Qiang
23
24 Kyoto
25
26 130 PU
27
28 SEE

0 tats, and author profiles for this publication... 1 2
1 0973
2 ransition Pattern Analysis for Travel Ro
3
4 Transactions on Information and Systems · Augu...
5
6 019EDP7096
7
8 READS
9
10 325
11
12
13
14 Sun Chenyi
15
16 University Kyoto
17
18 ICATIONS 22 CITATIONS 21 PUBL
19
20 PROFILE SEE P
21
22 Ma

Page: 2

0
0 VOL. E102-D
1
2 DECEMBER
3
4 The usage o
5
6 on Copyrigh
7
8 The author(
9
10 educational
11
12 Distribution

Page: 3

0
0
1 2472
2
3 PA PER
4 Us er Transition Pattern
.. ...
58 TheauthoriswiththeNationalInstituteofA
59
60 trial ScienceandTechnology(AIST),Tokyo,135-
61 a) E-mail:jj-sun@db.soc.i.kyoto-u.ac.jp
62 DOI:10.1587/transinf.2019EDP7096

2
0 IEICETRANS.INF.&SYST.,VOL.E102-D,N0.12DECEMBER...
1
2
3
4 Analysis for Travel Route Recommendation
.. ...
58 dvancedIndus-\nmetric, whereas they are symmet...
59
60 0064Japan.
61 gle; the green rectangle demonstrates the unba...
62 visitingordersbetweenlocations.

[63 rows x 3 columns]

Page: 4

0
0 SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
1
2
3 Fig.2 Vi sualization of
4 burgh;atra instationislo

```

..
62 city's implicit structure information, such as...
63 andtouristflow trends, andhelpstorestricttherou...
64 matched with the real situation. For instance,...
65 part of the visualized observed transition wei...
66 Edinburgh in the dataset (introduced in Sect.5...

```

```

...
st trip desig
onal resear
it. OP-var
aints have b
proposing s

```

```

2 3 4 5
0
1 2473
2
3 the observed transition w eight graph in Edin-
4 catedinthey ellowrectan gle. Thethi cknessof
..
62 n problem (TTDP) i s widely studied
63 ch, and OP is suitabl e as a the oretical
64 iant model s that sat isfy more realis-
65 een propos ed. These studies h ave fo-
66 uitable heu ristic plan ning algo rithms,

```

[67 rows x 6 columns]

Page: 5

```

0 1 \
0 2474
1
2 becauseitisanNP-hardtask. Forexample,teamOP(TOP) routes basedonly
3 extends OP by allowing multiple tours, and is ... we study location
4 multi-daytourplanning(e.g.,[7]). UnliketheTOP,... tion patterns to i
..
66 ...
67 wordmatching. i=1 j=2
68 However, the literature reviewed above rarely ...
69 ered location-location relations, and recommen... p -p +1\ ni j
70

```

```

2 3
0
1
2 onuse r-location matching. Inour work,
3 -locati on relations by considering transi-
4 mprove the existing tour recommendation
..
66 B\nij l 1 (4)
67
68
69 ≤(N- 1)(1-x ),∀i, j=2,⋯,N (5)\nij
70

```

[71 rows x 4 columns]

Page: 6

```

0 1
\
0 SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
1
2
3 |

```

4 where $R(p, p)$ is the reward function for visiti... gscorescanbeobta
 ..
 78 modelwithotherusers'travelroutedata. Forinstan... mtouristtravelrou
 79 rank support vector machine (RankSVM) is appli... betweentwoPOIs.
 80 byintroducingmorefeatures(e.g.,category,popula... aive idea to represe
 81 average visit duration) to rank POIs according... alize the observed
 82 query under the travel length budget constrain... sbetweenPOIsco

2 3
 0
 1 2475
 2
 3
 4 inedasPOI rewardbylearn-
 ..
 78 tedataanda ssignarewardto
 79
 80 nt transition pattern is to di-
 81 transitions. Then the transi-
 82 ldberegard edastherewards

[83 rows x 4 columns]
 Page: 7

0 \
 0
 1 2476
 2
 3 oftransitions. However,thisideahasseveraldrawb...
 4
 ..
 64 locations), we add an interaction latent compo...
 65 representlocalinteractionmodelsofthedecomposed...
 66 variables[25]. Hence,onereasonablesolutionfori...
 67 theweightedtransitionmatrixfromobserveddataist...
 68 torizetheobservedweightedtransitionmatrixT asf...

1
 0 IEICETRANS.INF.&SYST.,VOL.E102-D,NO.12DECEMBER...
 1
 2
 3
 4
 ..
 64 thehaversineformula[27](tocomputethedistancebe...
 65 twogeographiccoordinates).Second,wetakethereci...
 66 ofthevalueineachentrytocaptureaPOIdistancereci...
 67 calmatrix,inwhichthelongerthedistance,thesmall...
 68 value. Finally, wenormalizeeach entrybydividin...

[69 rows x 2 columns]
 Page: 8

0 \
 0
 1
 2 maximum value in the matrix, to construct the ...
 3 influence matrix $G \in \mathbb{R}^{N \times N}$. The spatial influen...

4 can be regarded as additional global knowledge...
 ..
 84 For every travel route of users, Fig.5 shows t...
 85 tribution of intervals between users' visiting...
 86 two POIs. These indicate how long tourists pre...
 87 travel from one POI to the next: most of the transits are
 88

0 1
 2477

1
 2
 3
 4

..
 84 the latent spatial feature V by solving the following...
 85 tive function:
 86 $(cid:6) (cid:6) \backslash n (cid:6) (cid:6) \backslash n^2$
 87 $\min (cid:6) E - V^T (cid:6) + \lambda (cid:9) V (cid:9)^2, \dots$
 88 V_e

[89 rows x 2 columns]
 Page: 9

0 \ 0
 2478

1
 2
 3
 4

..
 78
 79
 80

81 After assigning reward to locations and transitions...
 82 aim to combine knowledge of locations and tran...

1

0
 1
 2 recommend travel routes. Both location and tra...
 3 wards are already normalized and each value is...
 4 Therefore, we can combine the location and tra...

..
 78

79 To evaluate our proposed method with real tourists, we
 80

81 †Gurobi Optimization. <http://www.gurobi.com>
 82 †lp_solve package: <http://lpsolve.sourceforge.net>

[83 rows x 2 columns]
 Page: 10

0 1 2 3 \

0
 1
 2
 3 Table 1 Statistics of the Flickr image dataset.

4	Dataset #Photos #C	heck-ins #TravelRoutes #	Users
..
66	BytrainingaRankS	VMmodelwithPOIfeature	sunder
67			
68	the travel length bud	get, it recommends a trav	el route
69	basedonthePOIran	kingscore.	
70	• Markov. An expl	icit feature factorization	method

4
2479

0
1
2

3 which factorizes a pair of POIs by different t...

4 POIfeaturestocomputethetransitionprobabilitiesbe-

..

66 call T (how many of the user's real visited PO...

67

68 recommended) and tour precision T (how many re...

69 ommended POIs are in the user's real travel ro...

70 isappliedtoarecommendedtravelrouteexcludingthe

[71 rows x 5 columns]
Page: 11

0	1	2	3	\
0	Table2 Performancecompariso	noftravelroute	re	
1	termsoftourFlscore(TF1).			
2				
3	Budapest	Edinburgh		
4	PoiPop 0.216±0.277	0.331±0.378	0.3	
..	
74				
75	(cid:9)\n(cid:5)\n(T -T)2\nE			
76	RMSE = u U inf tru\n,\n U	(22)		
77				
78	Questionnairef	orRealTourists.Foreachu	serquery,	
	4		5	
0	commendation	undertheroutelengthbudgetin		
1				
2				
3	Glasgow	Toronto Vienna		
4	46±0.429	0.200±0.320	0.316±0.342	
..	
74			oi ank	
75	Altho\nRankSVM	ugh P R uses machine learning method\nwith mor...		
76	baselinePo	iPop,whichindicatesthatPOIpopularityismore		
77	efficienttha	notherfeatures. Forinstance,mosttouristswill		
78	notmissvi	sitingfamouslandmarksofthetargettravelcity.		

[79 rows x 6 columns]
Page: 12

0	1	\
0	Budapest Edinburgh	Glasgow
1	PoiPop 0.315±0.169 0.436±0.257	0.506±0.296
2	PoiRank 0.372±0.226 0.424±0.248	0.432±0.241
3	Markov 0.348±0.235 0.369±0.215	0.501±0.300

4 Tmf 0.292±0.204 0.423±0.246 0.418±0.253
 ..
 57 ducingtransitionknowledgetorecommendtravelroutes. Fig.8(a),fr
 58 ComparisonwithTransitionWeightInference. We erallandma
 59 use the RMSE of transition weights to evaluate... Forinstanc
 60 transition weight inference approaches, applie... corner, is a
 61 ure8(b) sh

	2	3	4	5	6
0	Toronto	Vienna			
1	0.386±0.201	0.400±0.222			
2	0.508±0.295	0.396±0.209			
3	0.454±0.277	0.321±0.181			
4	0.480±0.281	0.320±0.205			
..
57	omwhichwe	caneasily	findt	helocation	ofse
58	rksbecauset	heyhavem	oreli	nkstoothe	rPOI
59	e,FushimiIn	arishrine,l	ocated	inthelow	errig
60	famous sigh	tseeing att	ractio	n in Kyoto	. Fi
61	ows the infer	redtransiti	onwe	ight graph,	whic

[62 rows x 7 columns]
 Page: 13

0 \

0

1 2482

2

3 Fig.7 Impactofthetrade-offpara

4

5

6

7

8

9

10

11

12

13

14

15

16 Fig.8 Thevisualizationofobservedtransitionweig...

17 sitionweightsinthetourismdatasetofKyoto.

18

19

20

21 fillsunobservedtransitionweightsaccordingtothe...

22 transitions.

23 ImpactoftheTrade-offParameter. Asintroducedin

24 Sect.4.3,wecombinelocationandtransitionrewards...

25 ingatrade-offparameter α . Thisparametercontrols...

26 ativeweightsoflocationandlocationreward.Thelar...

27 valueof α ,themorethetourplanningdependsonlocati...

28

29 as opposed to transitions. Because this is a v...

30 parameterthatdirectlyaffectsthefinaltravelrout...

31 procedure, we have to discover the effects of ...

ues. Therefore, in order to understand the importance...
reward for each part, we evaluate the impact of α by plot...
recommendation performance under different α values.
Figure 7 plots the impact of the trade-off parameter in
terms of the recommendation performance F score...
the route length budget. We take the performance produced...
by PoiRank as the baseline that only uses location...
which will not change with the value of α . The...
types of transition knowledge produced by Markov...
41

1

0 IEICETRANS.INFO&SYST., VOL.E102-D, NO.12 DECEMBER...

1

2

3

meter α on different datasets.

4

5 GTmf, and TGTmf are included upon the PoiRank baseline.
6 According to the plot results, we observe that...
7 importance of location and transition varied in...
8 datasets (cities). This variation might demonstrate...
9 different travel styles of different cities. For...
10 transition patterns are more emphasized in Edinburgh...
11 now, according to Figs. 7(b) and 7(c), because...
12 performance decreases as α increases. This can also...
13 from Fig. 2, which shows that POI is in central Edinburgh...
14 tightly linked to each other. In contrast, cities...
15 like Taipei and Toronto rely more on location reward...
16 according to Figs. 7(a) and 7(d); the effects of transition...
17

18

these datasets are not obvious.

19 Another interesting finding from these figures...

20

ff

21 the trends of the different curves by introducing...
22 patterns are consistent with each other. The only difference...
23 might be caused by the inference ability of different...
24 datasets. For instance, all F scores in Figs. 7(a) and 7(d)...
25 usually increase as α rises, while the method Rank...
26 achieves the best result. The reason might be...
27 transition weight inference ability of our proposed...
28

mf

29 TGTmf, which includes additional spatial and temporal...

30

tures.

31 Finally, among all different values of α on different...
32 datasets, we find that a value between 0.5 and...
33 always achieves a higher performance upon the location...
34 reward. This could be an important hint for developing...
35 travel route recommendation systems with our proposed...
36 framework.

37 Evaluation of Real Tourists We collected feedback...
38 from seven foreign tourists according to the evaluation...
39 setting described in Sect. 5.2 and calculated simple...
40 for the results. The average rating scores of...
41

```

0
1
2
3
4
..
62
63
64 itingorderandsuggesteduserstotravelthroughNine...
65 andSanneizaka toKiyomizu-dera, whichiscorrecti...
66 itybecausetouristshavetoclimbthesepathsbeforee...

1
0 2483
1
2 Kiyomizu-dera.
3
4 6. Conclusion
..
62
63 Oudheusden,"Iteratedlocalsearchfortheteamorien...
64 lemwithtimewindows,"Computers&OperationsResear...
65 no.12,pp.3281-3290,2009.
66 [10] A.Garcia,P.Vansteenwegen,O.Arbelaitz,W.So...

[67 rows x 2 columns]
Page: 15
0 1 \

0
1 2484
2
3 Linaza,"Integratingpublictransportationinperso...
4 tourist guides," Computers & Operations Resear...
..
65 ofthearcorienteeringproblemmanditsapplicationto...
66 ning," Transportation research part E: logisti...
67 review,vol.68,pp.64-78,2014.
68 [30] R.Kohavietal., "Astudyofcross-validationandboot...
69 curacyestimationandmodelselection,"Appearsinth...

2
0 IEICETRANS.INF.&SYST.,VOL.E102-D,NO.12DECEMBER...
1
2
3 JointConferenceonArtificialIntelligence(IJCAI),1...
4 1145,Montreal,Canada,1995.
..
65
66
67
68
69

[70 rows x 3 columns]

```

vertical_strategy

Defines how to detect vertical lines / column boundaries. "text" → decide columns based on the positions of text. "lines" → use drawn vertical lines in the PDF.

"explicit" → use manually provided positions.

horizontal_strategy:

Defines how to detect horizontal lines / row boundaries. "text" → decide rows based on the y-coordinates of text. "lines" → use drawn horizontal lines in the PDF.

```
In [45]: tables= []

with pdfplumber.open('Final nm.pdf') as pdf:
    for page in pdf.pages:
        tables_on_page= page.extract_tables({
            'vertical_strategy': 'text',
            'horizontal_strategy': 'text',
            'intersection_x_tolerance': 10,
            'intersection_y_tolerance': 10
        })
        if tables_on_page:
            for table in tables_on_page:
                if table:
                    tables.append({
                        'page': pdf.pages.index(page) + 1,
                        'data': table
                    })

    for table in tables:
```

```
print("Page:", table['page'])  
print(pd.DataFrame(table['data']))
```

Page: 1

0 \
0 See discussions, s
1
2 User T
3
4 Article in IEICE
5
6 DOI: 10.1587/transinf.2
7
8 CITATIONS
9
10 7
11
12 3 authors:
13
14 Junjie
15
16 Kyoto
17
18 7 PUBL
19
20 SEE
21
22 Qiang
23
24 Kyoto
25
26 130 PU
27
28 SEE

0 tats, and author profiles for this publication... 1 2 0973
1
2 ransition Pattern Analysis for Travel Ro
3
4 Transactions on Information and Systems · Augu...
5
6 019EDP7096
7
8 READS
9
10 325
11
12
13
14 Sun Chenyi
15
16 University Kyoto
17
18 ICATIONS 22 CITATIONS 21 PUBL
19
20 PROFILE SEE P
21
22 Ma

23
24 Institute of Technology
25
26 BLICATIONS 1,053 CITATIONS
27
28 PROFILE

Page: 2

0
0 VOL. E102-D
1
2 DECEMBER
3
4 The usage o
5
6 on Copyrigh
7
8 The author(
9
10 educational
11
12 Distribution

Page: 3

0
0
1 2472
2
3 PA PER
4 Us er Transition Pattern
.. ...
58 TheauthoriswiththeNationalInstituteofA
59
60 trial ScienceandTechnology(AIST),Tokyo,135-
61 a) E-mail:jj-sun@db.soc.i.kyoto-u.ac.jp
62 DOI:10.1587/transinf.2019EDP7096

2
0 IEICETRANS.INF.&SYST.,VOL.E102-D,N0.12DECEMBER...
1
2
3
4 Analysis for Travel Route Recommendation
.. ...
58 dvancedIndus-\nmetric, whereas they are symmet...
59
60 0064Japan.
61 gle; the green rectangle demonstrates the unba...
62 visitingordersbetweenlocations.

[63 rows x 3 columns]

Page: 4

0
0 SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
1
2
3 Fig.2 Vi sualization of
4 burgh;atra instationislo

```

..
62 city's implicit structure information, such as...
63 andtouristflowtrends, andhelpstorestricttherou...
64 matched with the real situation. For instance,...
65 part of the visualized observed transition wei...
66 Edinburgh in the dataset (introduced in Sect.5...

```

```

...
st trip desig
onal resear
it. OP-var
aints have b
proposing s

```

```

2 3 4 5
0
1 2473
2
3 the observed transition w eight graph in Edin-
4 catedinthey ellowrectan gle. Thethi cknessof
..
62 n problem (TTDP) i s widely studied
63 ch, and OP is suitabl e as a the oretical
64 iant model s that sat isfy more realis-
65 een propos ed. These studies h ave fo-
66 uitable heu ristic plan ning algo rithms,

```

[67 rows x 6 columns]

Page: 5

```

0 0 1 \
1 IEICETRAN
2 2474
3 becauseitisanNP-hardtask. Forexample,teamOP(TOP) routes basedonly
4 extends OP by allowing multiple tours, and is ... we study location
..
67 ... x
68 wordmatching. i=1 j=2
69 However, the literature reviewed above rarely ...
70 ered location-location relations, and recommen... p -p +1\ni j
71

```

```

2 3
0 S.INF.& SYST.,VOL.E102-D,N0.12DECEMBER2019
1
2
3 onuse r-location matching. Inour work,
4 -locati on relations by considering transi-
..
67 B\nij l 1 (4)
68
69
70 ≤(N- 1)(1-x ),∀i, j=2,⋯ ,N (5)\nij
71

```

[72 rows x 4 columns]

Page: 6

```

0 0 1
\ SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
1
2
3 |

```

4 where $R(p, p)$ is the reward function for visiti... gscorescanbeobta
 ..
 78 modelwithotherusers'travelroutedata. Forinstan... mtouristtravelrou
 79 rank support vector machine (RankSVM) is appli... betweentwoPOIs.
 80 byintroducingmorefeatures(e.g.,category,popula... aive idea to represe
 81 average visit duration) to rank POIs according... alize the observed
 82 query under the travel length budget constrain... sbetweenPOIsco

2 3
 0
 1 2475
 2
 3
 4 inedasPOI rewardbylearn-
 ..
 78 tedataanda ssignarewardto
 79
 80 nt transition pattern is to di-
 81 transitions. Then the transi-
 82 ldberegard edastherewards

[83 rows x 4 columns]
 Page: 7

0 \
 0
 1 2476
 2
 3 oftransitions. However,thisideahasseveraldrawb...
 4
 ..
 64 locations), we add an interaction latent compo...
 65 representlocalinteractionmodelsofthedecomposed...
 66 variables[25]. Hence,onereasonablesolutionfori...
 67 theweightedtransitionmatrixfromobserveddataist...
 68 torizetheobservedweightedtransitionmatrixT asf...

1
 0 IEICETRANS.INF.&SYST.,VOL.E102-D,NO.12DECEMBER...
 1
 2
 3
 4
 ..
 64 thehaversineformula[27](tocomputethedistancebe...
 65 twogeographiccoordinates).Second,wetakethereci...
 66 ofthevalueineachentrytocaptureaPOIdistancereci...
 67 calmatrix,inwhichthelongerthedistance,thesmall...
 68 value. Finally, wenormalizeeach entrybydividin...

[69 rows x 2 columns]
 Page: 8

0 \
 0 SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
 1
 2
 3 maximum value in the matrix, to construct the ...

```

4   influence matrix  $G \in \mathbb{R}^{N \times N}$ . The spatial influen...
..
85  For every travel route of users, Fig.5 shows t...
86  tribution of intervals between users' visiting...
87  two POIs. These indicate how long tourists pre...
88  travel from one POI to the next: most of the transits are
89

```

```

0                                     1
                                     ENDATION
1                                     2477
2
3
4

```

```

..
85  the latent spatial feature  $V$  by solving the following...
86  tive function:
87  (cid:6) (cid:6) \n (cid:6) (cid:6) \n 2
88  min (cid:6)  $E - V^T (cid:6) + \lambda (cid:9) V (cid:9)^2$ , ...
89   $V_e$ 

```

[90 rows x 2 columns]
Page: 9

```

0                                     0 \
1                                     2478
2
3
4

```

```

..
79
80
81
82  After assigning reward to locations and transitions...
83  aim to combine knowledge of locations and tran...

```

```

1
0  IEICETRANS.INFO&SYST.,VOL.E102-D,N0.12DECEMBER...
1
2
3  recommend travel routes. Both location and tra...
4  wards are already normalized and each value is...
..
79
80  To evaluate our proposed method with real tourists, we
81
82  †Gurobi Optimization. http://www.gurobi.com
83  †lp_solve package: http://lpsolve.sourceforge.net

```

[84 rows x 2 columns]
Page: 10

```

0                                     1                                     2                                     3 \
0  SU  Neta l.: USER TRANSITION P  ATTERN ANALYSIS FOR TRAVEL  ROUTE RECOMM
1
2
3

```

4 Table1 Statistic softheFlickrimagedataset.

 67 BytrainingaRankS VMmodelwithPOIfeature sunder
 68
 69 the travel length bud get, it recommends a trav el route
 70 basedonthePOIran kingscore.
 71 • Markov. An expl icit feature factorization method

0 4
 ENDATION
 1 2479
 2
 3

4 which factorizes a pair of POIs by different t...
 ..
 67 call T (how many of the user's real visited PO...
 68
 69 recommended) and tour precision T (how many re...
 70 ommended POIs are in the user's real travel ro...
 71 isappliedtoarecommendedtravelrouteexcludingthe

[72 rows x 5 columns]
 Page: 11

0	1	2	3	\
0	Table2 Performancecompariso	noftravelroute	re	
1	termsoftourFlscore(TF1).			
2				
3	Budapest	Edinburgh		
4	PoiPop 0.216±0.277	0.331±0.378	0.3	
..	
74				
75	(cid:9)\n(cid:5)\n(T -T)2\nE			
76	RMSE = u U inf tru\n,\n U		(22)	
77				
78	Questionnairef	orRealTourists.Foreachu	serquery,	
0	4		5	
1	commendation	undertheroutelengthbudgetin		
2				
3	Glasgow	Toronto Vienna		
4	46±0.429	0.200±0.320	0.316±0.342	
..	
74			oi ank	
75	Altho\nRankSVM	ugh P R uses machine learning method\nwith mor...		
76	baselinePo	iPop,whichindicatesthatPOIpopularityismore		
77	efficienttha	notherfeatures. Forinstance,mosttouristswill		
78	notmissvi	sitingfamouslandmarksofthetargettravelcity.		

[79 rows x 6 columns]
 Page: 12

0	1	\
0		
1	Budapest Edinburgh	Glasgow
2	PoiPop 0.315±0.169 0.436±0.257	0.506±0.296
3	PoiRank 0.372±0.226 0.424±0.248	0.432±0.241

4 Markov 0.348±0.235 0.369±0.215 0.501±0.300
 ..
 58 ducing transition knowledge to recommend travel routes. Fig.8(a), fr
 59 Comparison with Transition Weight Inference. We erallandma
 60 use the RMSE of transition weights to evaluate... Forinstanc
 61 transition weight inference approaches, applie... corner, is a
 62 ure8(b) sh

	2	3	4	5	6
0					
1	Toronto	Vienna			
2	0.386±0.201	0.400±0.222			
3	0.508±0.295	0.396±0.209			
4	0.454±0.277	0.321±0.181			
..
58	om which we	can easily	find t	he location	of se
59	rks because t	hey have m	ore li	nkstoo the	rPOI
60	e, Fushimi In	arishrine, l	ocated	in the low	errig
61	famous sigh	tseeing att	ractio	n in Kyoto	. Fi
62	ows the infer	red transiti	on we	ight graph,	whic

[63 rows x 7 columns]
 Page: 13

0 \

0

1 2482

Fig.7 Impact of the trade-off para

Fig.8 The visualization of observed transitionweig...
 sition weights in the tourism dataset of Kyoto.

fills unobserved transition weights according to the...
 transitions.
 Impact of the Trade-off Parameter. As introduced in
 Sect.4.3, we combine location and transition rewards...
 in a trade-off parameter α . This parameter controls...
 a tive weights of location and location reward. The lar...
 value of α , the more the tour planning depends on locati...
 as opposed to transitions. Because this is a v...
 parameter that directly affects the final travel rout...
 procedure, we have to discover the effects of ...

ues. Therefore, in order to understand the importance...
reward for each part, we evaluate the impact of α by plot...
recommendation performance under different α values.
Figure 7 plots the impact of the trade-off parameter in
terms of the recommendation performance F score...
the route length budget. We take the performance prod...
by PoiRank as the baseline that only uses loca...
which will not change with the value of α . The...
types of transition knowledge produced by Mark...
41

1

0 IEICETRANS.INF.&SYST.,VOL.E102-D,N0.12DECEMBER...

1

2

3

meter α on different datasets.

4

5 GTmf, and TGTmf are included upon the PoiRank baseline.
6 According to the plot results, we observe that...
7 importance of location and transition varied i...
8 datasets (cities). This variation might demons...
9 ferent travel styles of different cities. For ...
10 tion patterns are more emphasized in Edinburgh...
11 gow, according to Figs. 7(b) and 7(c), because ...
12 mance decreases as α increases. This can also ...
13 from Fig. 2, which shows that POI is in central Edinburgh are
14 tightly linked to each other. In contrast, cit...
15 dapest and Toronto rely more on location rewar...
16 ing to Figs. 7(a) and 7(d); the effects of transition re...
17

18

these datasets are not obvious.

19 Another interesting finding from these figures...

20

ff

21 the trends of the different curves by introduci...
22 patterns are consistent with each other. The only dif...
23 might be caused by the inference ability of di...
24 ods. For instance, all F scores in Figs. 7(a) and 7(d)...
25 ually increase as α rises, while the method Ra...
26 achieves the best result. The reason might be ...
27 transition weight inference ability of our pro...
28

mf

29 TGT, which includes additional spatial and temporal...

30

tures.

31 Finally, among all different values of α on di...
32 datasets, we find that a value between 0.5 and...
33 ways achieve a higher performance upon the loc...
34 ward. This could be an important hint for deve...
35 travel route recommendation systems with our p...
36 framework.

37 Evaluation of Real Tourists We collected feedback
38 from seven foreign tourists according to the e...
39 setting described in Sect. 5.2 and calculated simple...
40 for the results. The average rating scores of ...
41

```

0  SUNetal.:USERTRANSITIONPATTERNANALYSISFORTRAVE...
1
2
3
4
..
63
64
65  itingorderandsuggesteduserstotravelthroughNine...
66  andSanneizaka toKiyomizu-dera, whichiscorrecti...
67  itybecausetouristshavetoclimbthesepathsbeforee...

```

```

                                1
0                                ENDATION
1                                2483
2
3                                Kiyomizu-dera.
4
..                                ...
63
64  Oudheusden,"Iteratedlocalsearchfortheteamorien...
65  lemwithtimewindows,"Computers&OperationsResear...
66                                no.12,pp.3281-3290,2009.
67  [10] A.Garcia,P.Vansteenwegen,O.Arbelaitz,W.So...

```

[68 rows x 2 columns]

Page: 15

```

0                                0                                1  \
1  2484
2
3  Linaza,"Integratingpublictransportationinperso...
4  tourist guides," Computers & Operations Resear...
..  ...                                ...
65  ofthearcorienteeringproblemmanditsapplicationto...
66  ning," Transportation research part E: logisti...
67  review,vol.68,pp.64-78,2014.
68  [30] R.Kohavietal., "Astudyofcross-validationandboot...
69  curacyestimationandmodelselection,"Appearsinth...

```

```

                                2
0  IEICETRANS.INF.&SYST.,VOL.E102-D,NO.12DECEMBER...
1
2
3  JointConferenceonArtificialIntelligence(IJCAI),1...
4  1145, Montreal, Canada, 1995.
..
65
66
67
68
69

```

[70 rows x 3 columns]

When pdfplumber tries to extract a table, it looks for the points where vertical and horizontal lines (or text boundaries) intersect → those intersections define table cells.

Without tolerance → strict, may miss cells.

With tolerance → flexible, captures more complete tables.

More forgiving → can detect tables even if lines/text are slightly misaligned(with the help of adding intersection)

any type of link there so we extract the links using this

```
In [46]: import fitz

doc= fitz.open('Final nm.pdf')
page= doc[0]

#get all link from the page:
links= page.links()
for link in links:
    print(link)
```

{'kind': 2, 'xref': 28, 'from': Rect(200.76278686523438, 71.55902099609375, 335.2672424316406, 79.510009765625), 'uri': 'https://www.researchgate.net/publication/335060973_User_Transition_Pattern_Analysis_for_Travel_Route_Recommendation?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_2&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 29, 'from': Rect(39.755008697509766, 90.77392578125, 55.2449951171875, 111.97662353515625), 'uri': 'https://www.researchgate.net/publication/335060973_User_Transition_Pattern_Analysis_for_Travel_Route_Recommendation?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_3&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 30, 'from': Rect(492.96209716796875, 33.129150390625, 564.5211181640625, 45.71826171875), 'uri': 'https://www.researchgate.net/?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_1&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 31, 'from': Rect(39.755008697509766, 222.6280517578125, 59.63251495361328, 245.8184814453125), 'uri': 'https://www.researchgate.net/profile/Junjie-Sun-5?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_4&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 32, 'from': Rect(66.25834655761719, 222.6280517578125, 95.41202545166016, 231.24163818359375), 'uri': 'https://www.researchgate.net/profile/Junjie-Sun-5?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_5&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 33, 'from': Rect(66.25834655761719, 233.2293701171875, 111.31402587890625, 241.84295654296875), 'uri': 'https://www.researchgate.net/institution/Kyoto-University?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_6&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 34, 'from': Rect(66.25834655761719, 258.4075927734375, 109.3262710571289, 272.32177734375), 'uri': 'https://www.researchgate.net/profile/Junjie-Sun-5?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_7&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 35, 'from': Rect(297.4999694824219, 222.6280517578125, 317.37750244140625, 245.8184814453125), 'uri': 'https://www.researchgate.net/profile/Chenyi-Zhuang?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_4&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 36, 'from': Rect(324.0033264160156, 222.6280517578125, 365.7460632324219, 231.24163818359375), 'uri': 'https://www.researchgate.net/profile/Chenyi-Zhuang?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_5&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 37, 'from': Rect(324.0033264160156, 233.2293701171875, 369.0589904785156, 241.84295654296875), 'uri': 'https://www.researchgate.net/institution/Kyoto-University?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_6&esc=publicationCoverPdf', 'id': ''}

{'kind': 2, 'xref': 38, 'from': Rect(324.0033264160156, 258.4075927734375, 367.0712585449219, 272.32177734375), 'uri': 'https://www.researchgate.net/profile/Chenyi-Zhuang?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_7&esc=publicationCoverPdf', 'id': ''}

```
MQ%3D%3D&el=1_x_7&_esc=publicationCoverPdf', 'id': ''}
{'kind': 2, 'xref': 39, 'from': Rect(39.755008697509766, 285.573486328125, 5
9.63251495361328, 308.763916015625), 'uri': 'https://www.researchgate.net/pr
ofile/Qiang-Ma-45?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrich
Source=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExM
Q%3D%3D&el=1_x_4&_esc=publicationCoverPdf', 'id': ''}
{'kind': 2, 'xref': 40, 'from': Rect(66.25834655761719, 285.573486328125, 9
2.09910583496094, 294.18707275390625), 'uri': 'https://www.researchgate.net/
profile/Qiang-Ma-45?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enri
chSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjE
xMQ%3D%3D&el=1_x_5&_esc=publicationCoverPdf', 'id': ''}
{'kind': 2, 'xref': 41, 'from': Rect(66.25834655761719, 296.1748046875, 147.
09353637695312, 304.78839111328125), 'uri': 'https://www.researchgate.net/in
stitution/Kyoto_Institute_of_Technology?enrichId=rgreq-fb334007aee6a77884a9c
a5e461b40b5-XXX&enrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI
0MzJAMTU3NTQ0MzQxNjExMQ%3D%3D&el=1_x_6&_esc=publicationCoverPdf', 'id': ''}
{'kind': 2, 'xref': 42, 'from': Rect(66.25834655761719, 321.35302734375, 10
9.3262710571289, 335.2672424316406), 'uri': 'https://www.researchgate.net/pr
ofile/Qiang-Ma-45?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&enrich
Source=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQxNjExM
Q%3D%3D&el=1_x_7&_esc=publicationCoverPdf', 'id': ''}
{'kind': 2, 'xref': 43, 'from': Rect(175.58462524414062, 802.3886108398438,
204.73829650878906, 811.6647338867188), 'uri': 'https://www.researchgate.ne
t/profile/Junjie-Sun-5?enrichId=rgreq-fb334007aee6a77884a9ca5e461b40b5-XXX&e
nrichSource=Y292ZXJQYWdlOzMzNTA2MDk3MztBUzo4MzIyODQxNjQ4ODI0MzJAMTU3NTQ0MzQx
NjExMQ%3D%3D&el=1_x_10&_esc=publicationCoverPdf', 'id': ''}
```

In []:

In []:

In []:

In []: