IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 11, 2020, accepted April 22, 2020, date of publication April 29, 2020, date of current version May 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2990916

An Adaptive Genetic Algorithm for Personalized Itinerary Planning

PHATPICHA YOCHUM, LIANG CHANG, TIANLONG GU, AND MANLI ZHU
Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

Corresponding author: Liang Chang (changl@guet.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant U1711263, Grant U1811264, and Grant 61966009, and in part by the Natural Science Foundation of Guangxi Province under Grant 2018GXNSFDA281045.

ABSTRACT Traveling as a very popular leisure activity enjoyed by many people all over the world. Typically, tourists have different kinds of preferences about their itineraries, limited time budgets, unfamiliar with the wide range of Points-of-Interest (POIs) in a city, so that planning an itinerary is quite tedious, time-consuming, and challenging for them. In this paper, we propose an adaptive genetic algorithm for personalized itinerary planning for travelers to plan their itineraries better. Firstly, desired starting POIs (e.g., POIs that are close to their hotels) and destination POIs (e.g., POIs that are near train stations or airports) are considered in our approach. Secondly, we also take some general factors into account that travelers would consider in their preferences of an itinerary, which are mandatory POIs, the total number of POIs, the overall POI popularity, the overall cost, and the overall rating. Thirdly, we view this kind of recommendation task as a Multi-Objective Optimization problem, and we propose an adaptive genetic algorithm with the crossover and mutation probabilities (AGAM) for solving this problem to better find the best global solution. Fourthly, we allocate different weights to every factor which considered in our paper to generate a personalized itinerary recommendation for better meet many kinds of preferences of tourists. Finally, we compare our approach against baselines on real-world datasets which include six touristic cities, and the experimental results show that the AGAM achieves better recommendation performance in terms of the mandatory POIs, total POI visits, overall POI popularity, total travel time (including travel time and visit duration), overall cost, and overall rating.

INDEX TERMS Itinerary planning, itinerary recommendation, travel recommendation, recommendation system, genetic algorithm.

I. INTRODUCTION

Research in the field of tourism recommendation systems has been studied for a long time [1]. Web technologies have become an effective resource for tourists. In addition, an increasing number of travel platforms are eager to build and share experiences and reviews of places, restaurants, and hotels. These contents and services prove useful for planning trips. However, common resources do not exhaustively cover a wide range of aspects. Users who seek various information thus need to search different kinds of sites and identify the most relevant information, which time-consuming and tedious.

At the same time, the rise of smartphone photography has resulted in plenty of photos being shared on the Web. Users often upload photos and share the location, emotion, status, etc. during their trips to others. These photos reflect their travel, activities, movements, and trajectories such as the visited place and the spent time there. Consequently, it is an outstanding way to understand itineraries about how to study the photo streams of tourists in the tourism area. Generally speaking, itinerary is a planned route or journey which consists of a tour with one or more travel destination choices. Most users have multiple choice for travel destinations, then choose Points-of-Interest (POIs), which suits their desires and budgets. The next step is to plan an itinerary by sorting POIs and a route as well as timetable arrangement. The goal of recommending an itinerary is to provide a sequence of visiting POIs, which must be finished during a limited time and covering the total cost of a trip.

Recent works have shown the effectiveness of geo-tagged photos in improving the itinerary recommendation

The associate editor coordinating the review of this manuscript and approving it for publication was Haiquan Zhao.

performance [2]–[5]. In particular, the main idea of these approaches is to learn a sequence of POIs and consider many factors such as user interest [6], POI popularity [8], POI category [9], and trip constraints like time [7] and cost [15] for constructing the itinerary planning models. However, most of these works are proposed based on the Orienteering Problem (OP) or traveling salesman problem (TSP) variants. Besides, they usually indicate the attraction of a single POI and consider user interest based on the review rating or the number of times a user has visited a place. The itinerary planning problems have been proved to be NP-hard and challenging. This is why the evolution approaches such as Genetic Algorithms (GA) have received increasing attention in this area recently [16]–[20]. The previous works on the itinerary recommendations use GA to solve such search and optimization problems. The robustness of GA is due to its capacity to locate the optimum in a multimodal landscape. Even GA is a powerful stochastic optimizer specialized in planning. Nevertheless, it still exhibits the shortcomings of the fixed crossover and mutation probabilities.

To address gaps in existing works, we formulate our personalized itinerary recommendation problem as a Multi-Objective Optimization (MOO) problem. An itinerary is defined as a path, which connected by specified starting POI and ending POI, and at least one other POI is contained. A good personalized itinerary planning is to recommend a tour that contains as many mandatory POIs and other POIs as possible, makes user's visit duration maximized within a time budget, makes overall popularity and overall rating maximized, and keeps user's cost on POI entrance as less as possible. Hence, mandatory POIs are the term of the most popular and special POIs that a tourist must be visited for a successful tour. We also determine the measure of the popularity of a POI from the average photo frequency at each POI and the measures of the user interest and the visit duration of all POIs from the average photo frequency of the user by leveraging the large collection of geotagged photos available online.

We propose a novel approach named AGAM, which is based on an adaptive genetic algorithm with the crossover and mutation probabilities, for solving our personalized itinerary recommendation problem. The factors for consideration should be given not only to the cost budget, time limitations, and starting/ending POIs, but also regards all the factors that tourists are interested in. All the above factors including mandatory POIs, the distance between POIs, and the length of the tour are considered to support tourists' demands. Especially, the popularity of POIs and user interest by the user's photo average photo frequencies from the Flickr photos, the rating of POIs in the sight reviews from the TripAdvisor, the time/distance between POIs from the Google Maps, and POIs list from the Wikipedia, these factors convey useful information regarding users' interests and habits. We try to make realistic itineraries and satisfy the tourists' demands as much as possible.

The main contributions of this paper can be summarized as the following:

- Firstly, we formulate the personalized itinerary recommendation task as the MOO problem and propose the AGAM approach to solve this problem using multi-consideration. We consider the mandatary POIs, the popularity of POIs, the rating of POIs, the starting/ending POIs, the distance between POIs, the cost budget, the time limitations, and the length of the tour to construct the itinerary planning, which suites to tourists demands. We also allocate different weights to every factor to generate a personalized itinerary recommendation for better meet many kinds of preferences of tourists.
- Secondly, we use the advantage of the large collection of geotagged photos to determine the travel history. Thus, the POI popularity, user interest, visit duration are measured by the users' photo average photo frequencies. Besides, we collect the rating of POI from the travelogue website, which can better reflect the popularity of POIs in the actual situation.
- Thirdly, we use real-world datasets on the Yahoo Flickr Creative Commons 100 Million Dataset (YFCC100M), Wikipedia.com, TripAdvisor.com, and Google Maps for six touristic cities. The results show that the AGAM outperforms better than baseline methods in terms of the mandatory POIs, total POI visits, overall POI popularity, total travel time (including travel time and visit duration), overall cost, and overall rating.

The remainder of this paper is organized as follows. Section II describes related work about travel itinerary recommendation systems. The overall framework is presented in Section III. Section IV introduces the preliminaries and problem definition. An adaptive genetic algorithm for itinerary recommendation is presented in Section V. Section VI analyzes the experimental results. The conclusion is provided in Section VII.

II. RELATED WORK

In recent years, the travel itinerary recommendation system is a highly topic in relevant specialized fields as computer science, operations research, and graph theory. It has attracted many researchers to design and plan a great travel itinerary and improve the accuracy of recommendations. An optimization model is a solution to solve the itinerary recommendation problems for getting the best itinerary planning. Many itinerary recommendation works are based on the Orienteering problem. For instance, Lim *et al.* [2] studied itinerary recommendation problems based on the Orienteering problem, where suggesting POIs according to user preferences and POI popularity. While the distance and travel time between POIs are minimized. Bolzoni *et al.* [32] utilized POI categories for solving the Orienteering problem with maximal category constraints and generating an optimal itinerary. Padia *et al.* [10] maximized the total profit from visiting POIs while the itinerary planning can be completed within a time budget.

Similarly, the optimization objective in [5] is for recommending itineraries including must-see POIs by maximizing the total score from visiting POIs under a fixed travel time budget. Lim *et al.* [7] considered time limitation in terms of queuing time. The Orienteering problem used to maximize both POI popularity and user interests and minimize queuing times.

In contrast to works based on the Orienteering problem, there are several optimization models like Heuristic optimization, Mathematical model, and other combinatorial optimization for solving multi-objective optimization problems. In [19], the objective function was minimized by an optimized GA during the search process while the high scores were reorganized in cross mutation phase. Based on user interests and trip constraints, Liu *et al.* [18] applied GA to the real-time route recommendation system by reducing the traffic jams and queuing time in POIs. Other like [20], the objective was to maximize the total scores in each POIs while maintaining the total travel time under constraints by GA. Wang *et al.* [11] extend the Ant Colony Optimization algorithm by merging user interests with POI popularity and using crowd data to recommend trips. Chang *et al.* [34] used a Greedy algorithm to minimize the process of trip planning and maximize user satisfaction with the best entertainment places while traveling to the destination.

Bolzoni *et al.* [9] applied a probabilistic algorithm to reduce the size of the input and get faster execution time for solving orienteering problems with category constraints. While Gaonkar *et al.* [4] used a reinforcement learning method to fit data by a probability distribution for the itinerary recommendation. A mathematical model was studied in [13] to maximize the risk-hedging ability and minimize the time budget. Jiaoman *et al.* [12] designed the itinerary arrangement with a mathematical model and solved the minimizing the total travel time under the total cost restriction. Mancinia and Steccab [36] developed a travel itinerary planning application by using a mathematical model to query data in a short computational time and using a mixed integer programming model to minimize operational costs under trip constraints.

The objective in [37] was minimizing the total cost on a trip in case of the entrance ticket of POIs, the cost for hotels, and the driving cost. Liu *et al.* [15] solved the problems of the popular route by considering the optimal route concatenation with the minimal travel cost. Cai *et al.* [33] presented itineraries in the short length while the popularity of POIs is optimized large in the travel time. Yu *et al.* [3] optimized the shortest path by finding a minimum spanning tree using the Prim algorithm. Rani *et al.* [35] determined the order of POI visits on a daily while the total travel distance was minimum. To optimized the minimum trip duration of each itinerary, Hsueh *et al.* [40] calculated the distances from the beginning location to the destination. Fogli and Sansonetti [6] exploited feedback from the user to maximize the user's satisfaction for the itinerary recommendation. Nurbakova *et al.* [14] merged the activities by maximizing the sum of the user's satisfaction scores within the itinerary under spatio-temporal constraints.

In the meantime, user preferences are a common factor in optimization objectives to obtain realistic travel trajectories in the model [21], [22], [31]. The maximize of the number of visited POIs and maximize the popular path were optimization in [23] by counting the frequent sequential patterns in the user's spatial and temporal behavior. Besides, Volkova *et al.* [25] added the category of POI to find user interest and Wörndl *et al.* [24] calculated the total score of the level of interest of POI by using a number of POIs per category. The user check-in behavior and visiting sequence were studied to construct transition time [26], [27]. The objectives were how to use the time for visiting POIs by maximizing the total time used on tour and minimizing the total distance traveled. They calculated the distance between POIs, visiting time, and travel time from both check-in data and geotagged photos. The task of public transportation was also considered as one factor of the environment. The optimization model with maximize travel comfort and convenience was established in [28]–[30]. Chen *et al.* [29] collected data from GPS device to predict route by the strategy of different modes of transportation.

III. OVERVIEW FRAMEWORK

In this section, we describe our overall itinerary recommendation framework as shown in Fig. 1. This framework composes of 4 steps: data collection, construction user travel history, calculation average photo frequency, and adaptive genetic algorithm for an itinerary recommendation.

Step 1: Data Collection First, we crawl a set of geotagged photos from Flickr and POIs list from Wikipedia. Second, we determine the POI visits in each city by date-time taken. Next, we map a photo to a POI by the latitude and the longitude coordinates. Besides, we retrieve additional detail like the rating of each POI from TripAdvisor and time/distance between POIs from Google Maps to meet various aspects of tourists.

Step 2: Construction User Travel History The next step after getting the POI visits is to construct the user travel history by sorting POI visits in ascending temporal order. Using the user travel history, we define each travel sequence by setting consecutive POI visits, in terms of the consecutive POI visits differ by 10 hours based on the ground truth of real-life user trajectories. Furthermore, we can measure the visit duration of POI based on the time between the first and last photo taken at each POI for each user.

Step 3: Calculation Average Photo Frequency From a set of the travel history of each user in Step 2, we can calculate the photo frequency. We calculate the average photo frequency at each POI to determine a measure of the popularity of a POI. Also, we calculate the average photo frequency of the user to determine a measure of the user interest and visit duration of all POIs.

Step 4: Adaptive Genetic Algorithm for Itinerary Recommendation We recommend a travel itinerary to the user based on an adaptive genetic algorithm. Our itinerary

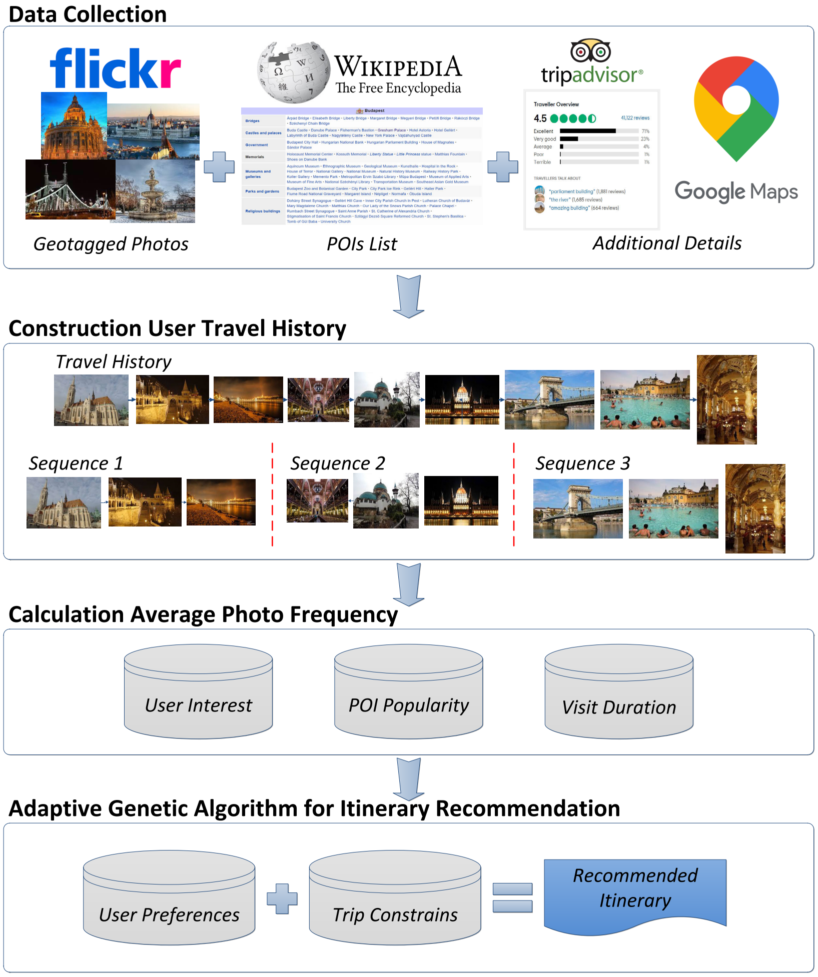


FIGURE 1. Overview of our proposed framework.

planning comprises a series of POIs and including mandatory POIs. We consider several factors such as the popularity of POIs, the rating of POIs, starting/ending POIs, the distance between POIs, the cost budget, time limitations, and the length of the tour to construct the itinerary planning.

IV. PRELIMINARIES AND PROBLEM DEFINITION

In this section, we first give some preliminaries used in our work, before formulating our itinerary recommendation problem.

A. PRELIMINARIES

We define preliminaries required to describe our approach as follows:

Definition 1 (Travel Trajectory): A user's travel trajectory is an ordered sequence, $Traj = ((q_1, a_{q_1}, d_{q_1}), \ldots, (q_k, a_{q_k}, d_{q_k}))$, where each triple (q_i, a_{q_i}, d_{q_i}) is comprised by a visited POI q_i, and the arrival time a_{q_i} and the departure time d_{q_i} at POI q_i. In addition, each POI q_i has two attributes lat_{q_i} and lng_{q_i}, which are its GPS coordinates.

Definition 2 (Travel Record): A user's travel record is the set $Reco(u_i) = ((q_1, fp_{q_1}, r_{q_1}), \ldots, (q_x, fp_{q_x}, r_{q_x}))$, where each triple (q_j, fp_{q_j}, r_{q_j}) is comprised by the visited POI q_j, the number of photos fp_{q_j} taken by user u_i, and the rating score r_{q_j} given by user u_i.

Definition 3 (Distance): We define the geospatial distance between two POIs as $Dist(q_i, q_j)$ which is obtained from Google Maps by using their GPS coordinates $lat_{q_i}, lng_{q_i}, lat_{q_j}$, and lng_{q_j}. Note that we only take the absolute value, therefore, $Dist(q_i, q_j)$ equals $Dist(q_j, q_i)$.

Definition 4 (POI Visit Duration): The visit duration of a POI is the average visit duration of all users who visited it. Let $U = u_1, \ldots, u_x$ be the set of the users of POI q_i, and the visit duration $Vdur(q_i)$ can be defined as follows:

$Vdur(q_i) = \frac{1}{x} \sum_{j=1}^{x} (d_{q_i}(u_j) - a_{q_i}(u_j))$

Definition 5 (Travel Time Between Two POIs): The travel time from q_i to q_j is $Trav(q_i, q_j) = Dist(q_i, q_j)/ps$, where ps is the travelling speed.

Definition 6 (POI Photo Frequency): Given the number of user's travel records x, a POI photo frequency can be represented as $Freq(q_i) = \sum_{j=1}^{x} fp_{qi}(u_j)$.

TABLE 1. Symbol definitions.

Symbol	Meaning
$MAXT$	The time budget
NOM	The number of mandatory POIs included in an itinerary q_i
NOP	The number of POIs included in an itinerary
$MAXP$	The largest popularity value among all POIs
$MAXV$	The longest visit duration value among all POIs
$MAXC$	The largest entrance cost among all POIs
$MAXR$	The largest rating value among all POIs
PC	The crossover probability
PM	The mutation probability

Definition 7 (POI Popularity): Here we use the photo frequency represents the popularity of a POI. As the photo frequency can be very large or very small, so we use $Popu(q_i) = 1/Fre(q_i)$ instead.

Definition 8 (POI Entrance Cost): A POI entrance cost $Cost(q_i)$ is the money you have to pay to enter the POI.

Definition 9 (POI Overall Rating): A POI overall rating is the average rating of all users', and it is represented as $Rati(q_i) = \frac{1}{x}\sum_{j=1}^{x} r_{q_i}(u_i)$.

B. PROBLEM DEFINITION

We now define our personalized itinerary recommendation problem as Multi-Objective Optimization (MOO) problem. In this paper, an itinerary is defined as a path, which connected by specified starting POI and ending POI, and at least one other POI is contained. In addition, all POIs in the itinerary can be visited only once.

Let $C = \{q_1, \ldots, q_L\}$ be the set of POIs, and $M = \{m_1, \ldots, m_K\}$ where $K < L$ be the set of mandatory POIs, ideally an itinerary can be represented as $I = \{q_s, \ldots, m_1, \ldots, m_K, \ldots q_d\}$, where q_s is the starting POI and q_d is the destination POI and $q_s, q_d \notin M$. The goal of personalized itinerary recommendation is to recommend a tour which contains as many mandatory POIs and other POIs as possible, makes user's visit duration maximized within a time budget, makes overall popularity and overall rating maximized, and keeps user's cost on POI entrance as less as possible.

V. ADAPTIVE GENETIC ALGORITHM FOR ITINERARY RECOMMENDATION

For solving the above problem, we present the AGAM with the crossover and mutation probabilities can be adjusted according to the fitness score during the iteration process. The following definitions introduce some symbols, so we can query the meaning of a symbol quickly using Table 1.

In our adaptive genetic algorithm, $P = \{p_1, \ldots, p_N\}$ is the set of population, where each individual $p_i = \{q_s, \ldots, q_d\}$ is a sequence of POIs, in another word, each single POI is a gene. Note that the gene sequence of each individual may have different lengths. As it is hard to measure different factors such as the inclusion of mandatory POIs, itinerary popularity and so on. Therefore, we use the normalization method to restrict all values within 1 to define our fitness function, which is shown as follows:

$$f(p_i) = w_1 Tm(p_i) + w_2 Tn(p_i) + w_3 Tp(p_i) + w_4 Tv(p_i) + w_5 Tc(p_i) + w_6 Tr(p_i) \quad (1)$$

where w_i is the weight of each factor, which can be adjusted to meet users' different preference, and $Tm(p_i)$ is the number of mandatory POIs included in p_i, $Tn(p_i)$ is the total number of POIs, $Tp(p_i)$ is the total popularity, $Tv(p_i)$ is the total popularity, $Tc(p_i)$ is the total entrance cost and $Tr(p_i)$ is the total rating. They are represented as follows:

$$Tm(p_i) = \frac{Nom(p_i)}{NOM} \quad (2)$$

$$Tn(p_i) = \frac{Nop(p_i)}{NOP} \quad (3)$$

$$Tp(p_i) = \frac{Totp(p_i)}{MAXP} \quad (4)$$

where $Totp(p_i) = Popu(q_s) + \ldots + Popu(q_d)$

$$Tv(p_i) = \frac{Totv(p_i)}{MAXV} \quad (5)$$

where $Totv(p_i) = Vdur(q_s) + \ldots + Vdur(q_d)$

$$Tc(p_i) = \frac{Totc(p_i)}{MAXC} \quad (6)$$

where $Totc(p_i) = Cost(q_s) + \ldots + Cost(q_d)$

$$Tr(p_i) = \frac{Totr(p_i)}{MAXR} \quad (7)$$

where $Totr(p_i) = Rati(q_s) + \ldots + Rati(q_d)$

The crossover and mutation probabilities are dynamic in our AGAM, and this is helpful for finding the best solution and preventing the program falls into the local best solution. They are defined as follows:

$$PC = \begin{cases} pc_1 - \dfrac{(pc_1 - pc_2)\left(f' - f_{avg}\right)}{f_{max} - f_{avg}}, & f' \geq f_{avg} \\ pc_1, & f' < f_{avg} \end{cases} \quad (8)$$

where pc_1 and pc_2 are parameters, f' is the larger fitness score of two individuals, which are going to generate next generation, f_{max} is the largest fitness score among a population, and f_{avg} is the average fitness score of a population.

$$PM = \begin{cases} pm_1 - \dfrac{(pm_1 - pm_2)\left(f_{max} - f\right)}{f_{max} - f_{avg}}, & f \geq f_{avg} \\ pm_1, & f < f_{avg} \end{cases} \quad (9)$$

where pm_1 and pm_2 are parameters, and f is the fitness score of an individual which is going to mutate.

We notice that some unvisited POIs are expected to be inserted to each individual of the next generation in order to improve the solution quality because extra time may be

Algorithm 1 AGAM Algorithm

Input: time budget $MAXT$,
population size N,
crossover rate pc_1, pc_2,
mutation rate pm_1, pm_2,
factors weight $w_1, w_2, w_3, w_4, w_5, w_6$,
iteration number δ

Output: best solution p_{opt}

generate N individuals randomly as initially population set P, with each individual within the time budget constraint
  TotalTimeCost(p_i) $\leq MAXT$ for each $p_i \in P$
for $i = 1$ **to** δ **do**
  for $j = 1$ **to** $N - 1$ **do**
    select two individuals p_j and p_{j+1} from P
    get PC by equation (8)
    generate p_a and p_b of p_j and p_{j+1} by one-point crossover and PC
    if TotalTimeCost(p_a) $> MAXT$ or $TotalTimeCost(p_b) > MAXT$ **then**
      regenerate p_a and p_b
    end if
    if TotalTimeCost(p_a) $< MAXT$ **then**
      try to insert unvisited POIs into p_a
    end if
    if TotalTimeCost(p_b) $< MAXT$ **then**
      try to insert unvisited POIs into p_b
    end if
    save p_c and p_d to P_1
  end for
  for $j = 1$ **to** N **do**
    select an individual p_j from P_1
    get PM by equation (9)
    randomly select a gene position from p_j and mutate it to generate a new individual p'_a by PM
    if TotalTimeCost(p'_a)$> MAXT$ **then**
      regerate p'_a
    end if
    if TotalTimeCost(p'_a) $< MAXT$ **then**
      try to insert unvisited POIs into p'_a
    end if
    update p_a with p'_a in P_1
  end for
  update $P = P_1$
end for
return the best solution p_{opt} in P

available after the operations of crossover and mutation. To this end, we try to insert some candidate POIs into every single individual of the next generation with the total travel time within the time budget.

It is worth to mention that, for each individual p_i, the total time cost $TotalTimeCost(p_i) = Tra(q_s, q_{s+1}) + \ldots + Trav(q_{d-1}, q_d) + Tv(p_i)$ is no more than the time budget $MAXT$.

TABLE 2. Statistics of these datasets.

Set ID	City	POIs	Users	POI Visits	Photos
1	Budapest	37	3,940	18,513	36,000
2	Edinburgh	28	4,844	33,944	82,060
3	Toronto	29	4,050	39,419	157,505
4	Glasgow	25	1,582	11,434	29,019
5	Perth	22	430	3643	18,462
6	Osaka	25	952	7747	392,420

In our AGAM, the parameters $pc_1 = 0.9, pc_2 = 0.6, pm_1 = 0.1, pm_2 = 0.001$ which are set based on the experience from common usage in dynamic crossover and mutation probabilities, and the factors weights used in the number of mandatory POIs, the number of POIs, overall popularity, overall visit duration, overall rating, overall cost are $w_1 = 10, w_2 = 5, w_3 = 2, w_4 = 2, w_5 = 1, w_6 = 1$ respectively. The larger weight values represent the more important factors. The whole process of our proposed AGAM is outlined in Algorithm 1 in detail.

VI. EXPERIMENTS

In this section, we describe our experiments, which include our datasets, baseline algorithms, evaluation metrics, and results and discussion.

A. DATASETS

For our experiment and analysis, we use real-world datasets from the Yahoo! Flickr Creative Commons 100M [41], which contains 100 million photos and videos. The list of POIs is collected from [42]–[47]. The rating of POIs in the sight reviews are crawled from the TripAdvisor. We count time/distance between POIs from the Google Maps. Table 2 lists statistics of these data sets. There are six cities: Budapest, Edinburgh, Toronto, Glasgow, Perth, and Osaka.

In our experiment, the starting and destination POIs, and mandatory POIs in each itinerary are generated randomly and recorded. In addition, we generate 300 these itineraries for each city.

B. BASELINE ALGORITHMS

We compare our proposed AGAM against different baseline algorithms to evaluate its recommendation performance.

1) **AGAM (our proposed model)** Firstly, generating N itineraries as initial population P. For each individual $p_i = p_{i1} \cup p_{i2}$, where p_{i1} is generated first, and it includes the start POI, the destination POI and mandatory POIs, p_{i2} is generated after p_{i1}, and it includes other POIs. Secondly, the AGA algorithm is used to generate the best solution (the recommended itinerary) which contains the most mandatory POIs and other POIs, user's visit duration is maximized within a time budget, overall popularity and overall rating are maximized, and user's cost on POI entrance is low. The model considers a personalized itinerary which generally includes popular or special POIs where tourists

often want to visit while other needs such as POI popularity and visit duration are also considered.

2) **MaxN** Generates an itinerary with the most POIs. Mandatory POIs are added first according to their visit duration time by the greedy strategy, then the same way is utilized to the remaining POIs. This approach focuses on the number of POIs included in an itinerary only without considering any other factors, so an itinerary that has the most mandatory POs and the most POIs within the time budget is much preferred. It provides the baseline in the number of mandatory POIs and the number of POIs.
3) **MaxP** Generates an itinerary with a relatively large overall popularity. Mandatory POIs are added first according to their popularity value by the greedy strategy, then the same way is utilized to the other POIs. This approach focuses on the popularity factor only without considering any other factors, so an itinerary that has a large overall popularity value is much preferred. It provides the popularity factor baseline in our paper.

The algorithms used for this work were implemented using the C++ programming language.

C. EVALUATION METRICS

We evaluate the performance of our algorithm and the baselines, which involve recommending itinerary planning. Our algorithm is based on an adaptive genetic algorithm to recommend a tour which contains as many mandatory POIs and other POIs as possible, makes user's visit duration maximized within a time budget, makes overall popularity and overall rating maximized, and keeps user's cost on POI entrance as less as possible. Thus, our algorithm utilizes evaluation metrics for the itinerary recommendation as follows:

1) **Mandatory POIs** The number of mandatory POIs included in the recommended itinerary.
2) **POI Visited** The number of POIs included in the recommended itinerary.
3) **Travel Time** The total travel time from one POI to another POI by the sequence in the recommended itinerary.
4) **Visit Duration** The sum of visit duration at each POI in the recommended itinerary.
5) **POI popularity** The value of the popularity of all POIs in the recommended itinerary.
6) **Cost** The total cost of all POIs in which entrance fee in the recommended itinerary.
7) **Rating** The total score of the rating of all POIs in the recommended itinerary.

D. RESULTS AND DISCUSSION

We show and discuss the experimental results in terms of mandatory POIs, POI visited, travel time, visit duration, the popularity of POIs, cost, and ratings. Furthermore, we considered four mandatory POI sets including one POI, two POIs, three POIs, and four POIs respectively, and they are randomly selected from the whole POI set of a city.

TABLE 3. The number of successful itineraries (out of 100) which included mandatory POIs. Higher values are better and the best figures among AGAM, MaxN, MaxP are in bold.

	AGAM mandatory POIs				MaxN mandatory POIs				MaxP mandatory POIs			
City	1	2	3	4	1	2	3	4	1	2	3	4
Budapest	**100**	**98**	**91**	**77**	**100**	**98**	90	76	**100**	**98**	90	75
Edinburgh	**95**	**94**	90	**65**	**95**	**94**	**92**	**65**	**95**	**94**	**92**	**65**
Toronto	**89**	69	37	5	**89**	**71**	**40**	8	**89**	**71**	**40**	**9**
Glasgow	**100**	**100**	**92**	56	**100**	**100**	89	**62**	**100**	**100**	89	**62**
Perth	**100**	**97**	**84**	62	**100**	**97**	83	62	**100**	**97**	83	**63**
Osaka	**96**	67	**31**	6	**96**	**68**	29	6	**96**	**68**	29	**8**

TABLE 4. Average number of POIs visited including failed itineraries by algorithm, mandatory POIs set size and city. Higher values are better and the best figures among AGAM, MaxN, MaxP are in bold.

	AGAM mandatory POIs				MaxN mandatory POIs				MaxP mandatory POIs			
City	1	2	3	4	1	2	3	4	1	2	3	4
Budapest	**7.8**	**7.5**	**7.1**	**7.1**	6.4	6.7	6.6	6.7	4.1	4.9	5.7	6.2
Edinburgh	**7.8**	**7.6**	**7.4**	**7.0**	6.2	6.1	6.4	6.6	7.3	7.5	7.2	6.8
Toronto	**5.5**	**5.4**	**5.2**	**5.1**	4.5	4.7	4.6	4.7	4.0	4.3	4.5	4.6
Glasgow	**7.2**	**6.7**	**6.5**	**6.4**	4.9	5.0	5.2	5.7	6.3	6.1	6.1	6.1
Perth	**8.2**	**7.7**	**7.5**	**7.3**	6.6	6.5	6.6	6.5	7.4	7.3	7.2	7.1
Osaka	**5.6**	**5.4**	**5.3**	**5.2**	3.6	4.0	4.3	4.5	3.9	4.0	4.3	4.6

1) TOTAL MANDATORY POIs VISITED IN RECOMMENDED ITINERARIES

We name an itinerary which includes every element in a mandatory set as a successful itinerary. The result of the number of successful itineraries generated by the AGAM, MaxN, and MaxP, along with the different sizes of the mandatory POI sets is shown in Table 3. The success rate of each method is nearly the same, and it clearly shows that the success rate is going down with the increase in the number of mandatory POIs. We can see that MaxN and MaxP achieve better performance, this is because both MaxN and MaxP add the mandatory POIs first then add the other POIs by the greedy strategy with considering different factors to generate the recommended itinerary, but our AGAM adds every single POI randomly and it still achieves a relatively good result.

2) TOTAL POI VISITS IN RECOMMENDED ITINERARIES

Table 4 presents the average number of total POIs visited of itineraries that every algorithm generated. Overall, the AGAM has the best performance among all algorithms over all the cities. The figures see a decline along with the increase of the mandatory POI set size, and the reason is that large mandatory POI sets restrict algorithms' ability of including more other POIs because some mandatory POIs may have long visit time or takes a long time to travel to another POI, so there is very limited time for other POIs. Besides, although MaxN generates its itineraries by priority adding POIs which

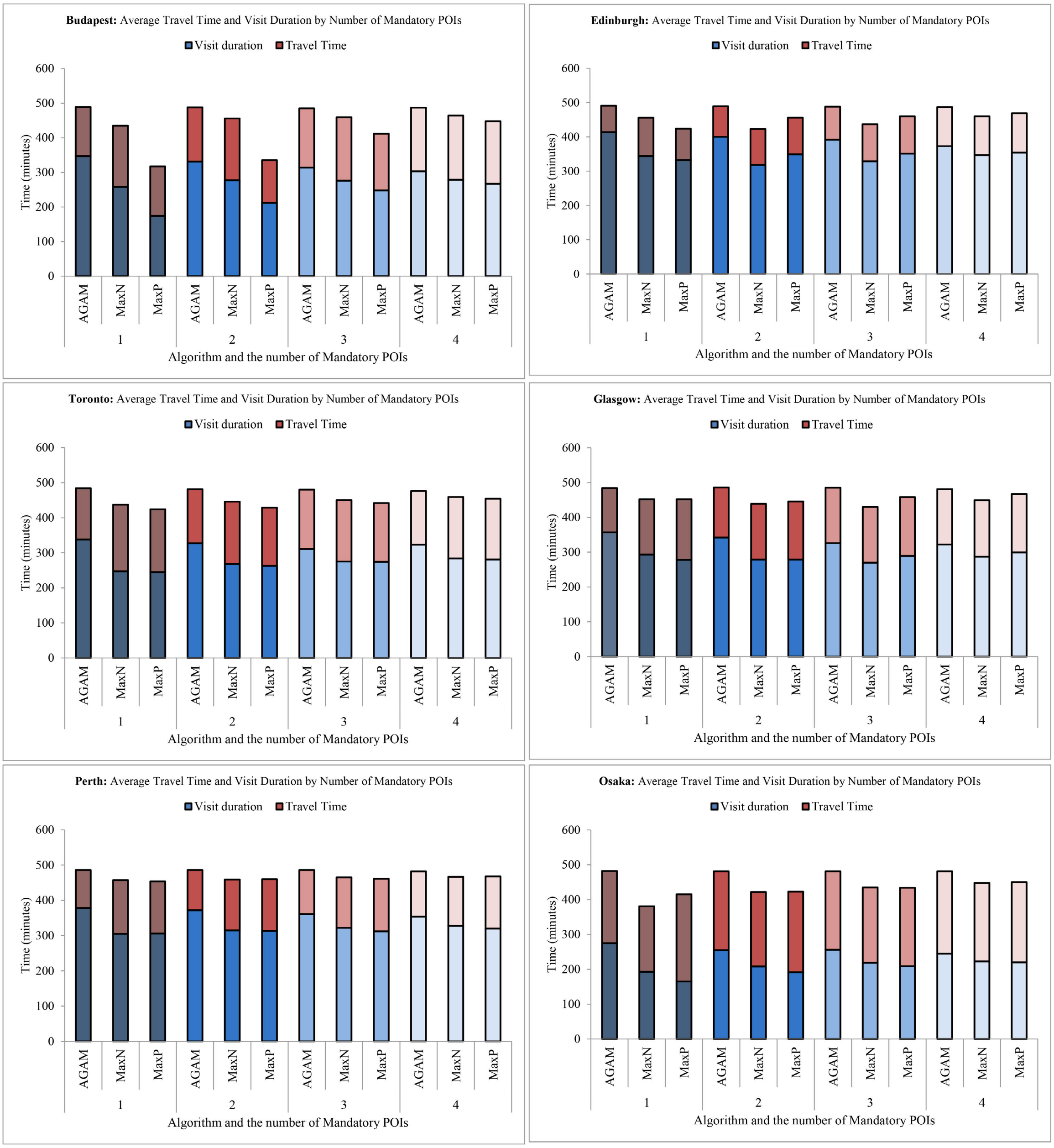


FIGURE 2. Average total travel time and visit duration, the total travel time is represented by top columns while the visit duration is represented by bottom columns. Lower values for travel time are better, while higher values for visit duration are better.

have short visit duration, the travel time between POIs is out of its control, so the results of MaxN are not always better than MaxP over all the cities.

3) TOTAL TRAVEL TIME AND VISIT DURATION IN RECOMMENDED ITINERARIES

As shown in Fig. 2, we can see that the total travel time for the itineraries generated by all methods was never exceeded the time budget. The AGAM uses the time budget efficiently in all six cities, in the meanwhile, the total visit duration is also better than MaxN and MaxP. In addition, for every method, we can see that the figures for all mandatory POI sets are similar to each other, this is because we have a hard constraint on travel time, which is time budget.

4) TOTAL POI POPULARITY IN RECOMMENDED ITINERARIES

The results of the total POI popularity of recommended tours are shown in Fig. 3. It is clear that the AGAM algorithm gets the highest values in every city, and the main reason is that we allocate a relatively high weight to the factor of POI popularity. Additionally, the MaxP archives better results than MaxN in most cities, and the reason is that MaxP generates its itineraries by priority adding POIs which have large popularity.

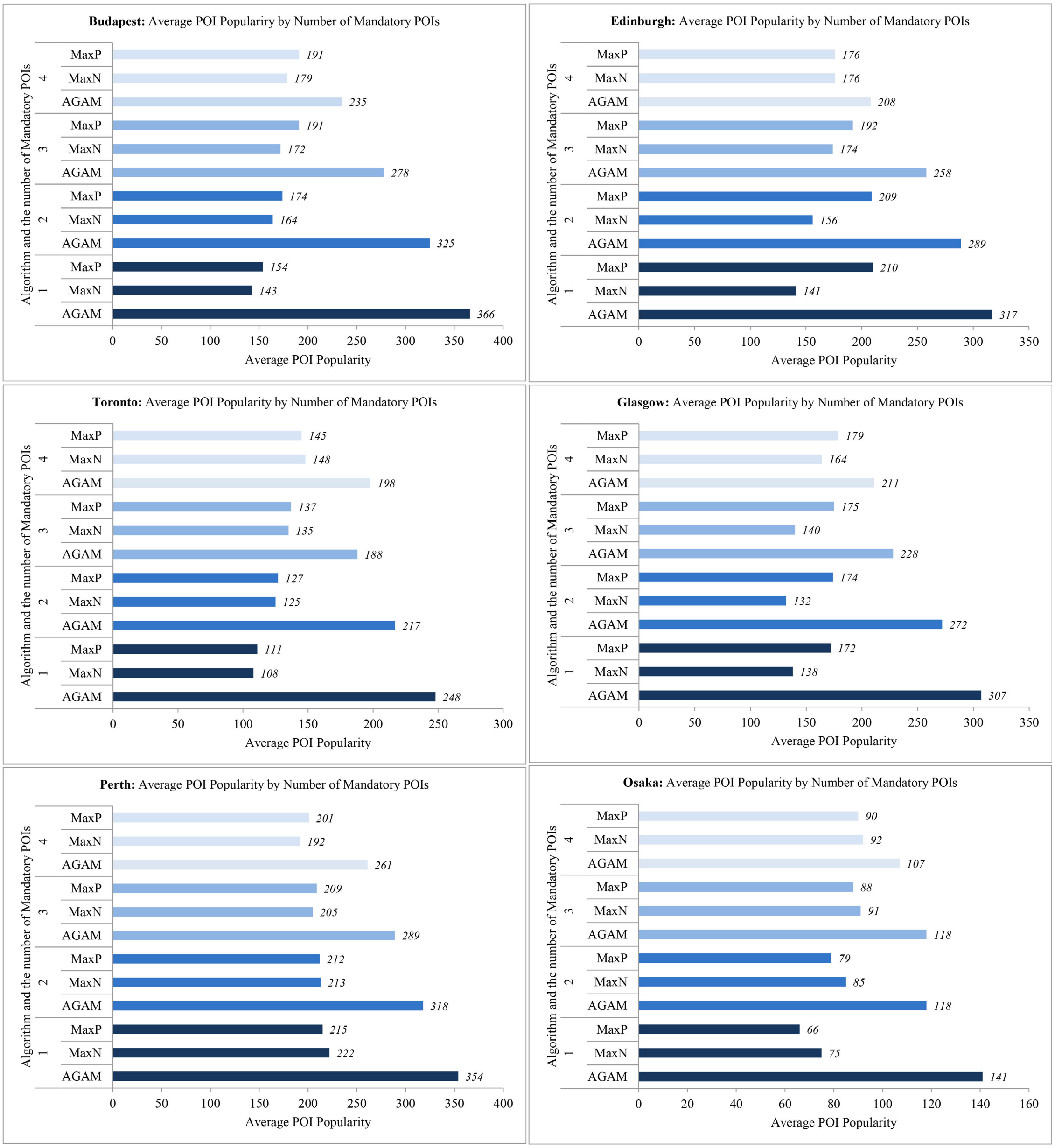


FIGURE 3. Average overall popularity of itineraries by mandatory POI set size. Higher values are better.

5) TOTAL COST IN RECOMMENDED ITINERARIES

As shown in Table 5, the MaxN and MaxP algorithms achieve better performance. This is reasonable because the AGAM includes more POIs than other algorithms as shown in Table 4. And in our experiments, we have not given a high weight to the factor of cost but a higher weight to the factor of the number of POI visited, so the total cost is relatively high.

6) TOTAL RATING IN RECOMMENDED ITINERARIES

Table 6 presents the total rating score of all POIs in the recommended itinerary. The results show that the AGAM generally outperforms all two baselines over all the cities in terms of rating. This is because the AGAM is able to generate itineraries which include more POIs than MaxN and MaxP as shown in Table 3, therefore, although we do not give a high weight to the factor of rating, it still can achieve a better result.

TABLE 5. Average overall cost of itineraries by algorithm, mandatory POIs set size and city. Smaller values are better and the best figures among AGAM, MaxN, MaxP are in bold.

	AGAM mandatory POIs				MaxN mandatory POIs				MaxP mandatory POIs			
City	1	2	3	4	1	2	3	4	1	2	3	4
Budapest	30	33	34	35	46	49	46	42	**23**	**28**	**31**	**33**
Edinburgh	**27**	**28**	**28**	**28**	39	37	35	32	32	35	37	35
Toronto	**36**	**38**	**40**	**38**	42	39	**40**	41	49	45	48	40
Glasgow	**3**	**4**	**5**	4	15	14	8	5	**3**	**4**	**5**	5
Perth	18	18	20	21	**11**	**13**	**16**	**17**	36	36	35	32
Osaka	20	20	21	20	19	20	19	**18**	**16**	**18**	**18**	**18**

TABLE 6. Average overall rating of itineraries by algorithm, mandatory POIs set size and city. Higher values are better and the best figures among AGAM, MaxN, MaxP are in bold.

	AGAM mandatory POIs				MaxN mandatory POIs				MaxP mandatory POIs			
City	1	2	3	4	1	2	3	4	1	2	3	4
Budapest	**33.1**	**32**	**30.4**	**30.2**	26.6	28.1	27.9	28.3	17.2	20.7	24	26.2
Edinburgh	**34.4**	**33.2**	**32.3**	**30.6**	26.8	26.3	27.7	28.7	31.1	32.7	31.2	29.7
Toronto	**23**	**22.6**	**21.9**	**21.5**	19.3	19.8	19.7	19.8	16.9	18	18.9	19.3
Glasgow	**29.3**	**27.9**	**27.3**	**26.8**	21.3	21.2	21.7	23.5	25.5	25.1	25.2	25.2
Perth	**33.2**	**31.3**	**30.5**	**29.5**	26.4	26.0	26.2	25.8	30	29.4	29	28.4
Osaka	**22.2**	**21.2**	**20.9**	**20.7**	14.5	15.9	17.3	18	15.3	16.1	17.3	18.2

VII. CONCLUSION

In this paper, we proposed an adaptive genetic algorithm for personalized itinerary planning for travelers. Firstly, desired starting POIs and destination POIs were considered in our approach. Secondly, we also took some general factors into account that travelers would consider in their preferences of an itinerary, which are mandatory POIs, the total number of POIs, the overall POI popularity, the overall cost, and the overall rating. Thirdly, we viewed this kind of recommendation task as a Multi-Objective Optimization problem, and we proposed the AGAM for solving this problem, which is based on an adaptive genetic algorithm with the crossover and mutation probabilities to better find the best global solution. Fourthly, we allocated different weights to every factor for generating the personalized itinerary planning to better meet many kinds of preferences of tourists. Finally, we compared our approach against baselines on real-world datasets which include six touristic cities, and the experimental results showed that the AGAM outperforms better than baseline methods in terms of the mandatory POIs, total POI visits, overall POI popularity, total travel time (including travel time and visit duration), overall cost, and overall rating.

We simply allocated a larger integer to each factor to represent the users' preference, a better weight allocation rule could be employed in the future. Also, we will consider the various modes of transport and sentiments such as opinions and reviews for user preferences and route planning.

REFERENCES

[1] P. Yochum, L. Chang, T. Gu, and M. Zhu, "Linked open data in location-based recommendation system on tourism domain: A survey," *IEEE Access*, vol. 8, pp. 16409–16439, 2020.

[2] K. H. Lim, J. Chan, C. Leckie, and S. Karunasekera, "Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency," *Knowl. Inf. Syst.*, vol. 54, no. 2, pp. 375–406, Feb. 2018.

[3] Y. Yu, Y. Zhao, G. Yu, and G. Wang, "Mining coterie patterns from Instagram photo trajectories for recommending popular travel routes," *Frontiers Comput. Sci.*, vol. 11, no. 6, pp. 1007–1022, Dec. 2017.

[4] R. Gaonkar, M. Tavakol, and U. Brefeld, "MDP-based itinerary recommendation using Geo-tagged social media," in *Proc. IDA*, 2018, pp. 111–123.

[5] K. Taylor, K. H. Lim, and J. Chan, "Travel itinerary recommendations with must-see points-of-interest," in *Proc. Companion Web Conf. Web Conf. (WWW)*, 2018, pp. 1198–1205.

[6] A. Fogli and G. Sansonetti, "Exploiting semantics for context-aware itinerary recommendation," *Pers. Ubiquitous Comput.*, vol. 23, no. 2, pp. 215–231, Apr. 2019.

[7] K. H. Lim, J. Chan, S. Karunasekera, and C. Leckie, "Personalized itinerary recommendation with queuing time awareness," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 325–334.

[8] X. Peng and Z. Huang, "A novel popular tourist attraction discovering approach based on geo-tagged social media big data," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 7, p. 216, 2017.

[9] P. Bolzoni, F. Persia, and S. Helmer, "Itinerary planning with category constraints using a probabilistic approach," in *Proc. DEXA*, vol. 2, 2017, pp. 363–377.

[10] P. Padia, B. Singhal, and K. H. Lim, "User-relative personalized tour recommendation," in *Proc. IUI Workshops*, 2019, pp. 1–6.

[11] X. Wang, C. Leckie, J. Chan, K. H. Lim, and T. Vaithianathan, "Improving personalized trip recommendation by avoiding crowds," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2016, pp. 25–34.

[12] D. Jiaoman, L. Lei, and L. Xiang, "Travel planning problem considering site selection and itinerary making," in *Proc. Conf. Res. Adapt. Convergent Syst. (RACS)*, 2018, pp. 29–36.

[13] Y. Zhang and J. Tang, "Itinerary planning with time budget for risk-averse travelers," *Eur. J. Oper. Res.*, vol. 267, no. 1, pp. 288–303, May 2018.

[14] D. Nurbakova, L. Laporte, S. Calabretto, and J. Gensel, "Itinerary recommendation for cruises: User study," in *Proc. RecTour RecSys*, 2017, pp. 31–34.

[15] H. Liu, C. Jin, and A. Zhou, "Popular route planning with travel cost estimation from trajectories," *Frontiers Comput. Sci.*, vol. 14, no. 1, pp. 191–207, Feb. 2020.

[16] W. Yongzhen, C. Yan, and Y. Yingying, "Improved grouping genetic algorithm for solving multi-traveling salesman problem," *J. Electron. Inf.*, vol. 39, no. 1, pp. 198–205, 2017.

[17] G. Kobeaga, M. Merino, and J. A. Lozano, "An efficient evolutionary algorithm for the orienteering problem," *Comput. Oper. Res.*, vol. 90, pp. 42–59, Feb. 2018.

[18] L. Liu, J. Xu, S. S. Liao, and H. Chen, "A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3409–3417, Jun. 2014.

[19] C. Yuan and M. Uehara, "Improvement of multi-purpose travel route recommendation system based on genetic algorithm," in *Proc. 7th Int. Symp. Comput. Netw. Workshops (CANDARW)*, Nov. 2019, pp. 305–308.

[20] B. S. Wibowo and M. Handayani, "A genetic algorithm for generating travel itinerary recommendation with restaurant selection," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2018, pp. 427–431.

[21] L. Hang, S.-H. Kang, W. Jin, and D.-H. Kim, "Design and implementation of an optimal travel route recommender system on big data for tourists in Jeju," *Processes*, vol. 6, no. 8, p. 133, 2018.

[22] J. Han and H. Lee, "Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media," *Pervas. Mobile Comput.*, vol. 18, pp. 4–17, Apr. 2015.

[23] C.-Y. Tsai and B.-H. Lai, "A location-item-time sequential pattern mining algorithm for route recommendation," *Knowl.-Based Syst.*, vol. 73, pp. 97–110, Jan. 2015.

[24] W. Wörndl, A. Hefele, and D. Herzog, "Recommending a sequence of interesting places for tourist trips," *Inf. Technol. Tourism*, vol. 17, no. 1, pp. 31–54, Mar. 2017.

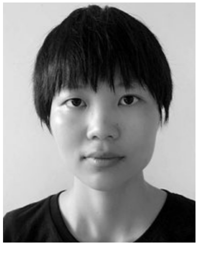
[25] L. Volkova, E. Yagunova, E. Pronoza, A. Maslennikova, D. Bliznuk, M. Tokareva, and A. Abdullaev, "Recommender system for tourist itineraries based on aspects extraction from reviews corpora," *Polibits*, vol. 57, pp. 81–88, Jan. 2018.

[26] X. Zhu, R. Hao, H. Chi, and X. Du, "FineRoute: Personalized and time-aware route recommendation based on check-ins," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10461–10469, Nov. 2017.
[27] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Trans. Hum.-Machine Syst.*, vol. 46, no. 1, pp. 151–158, Feb. 2016.
[28] M. Socharoentum and H. A. Karimi, "Multi-modal transportation with multi-criteria walking (MMT-MCW): Personalized route recommender," *Comput., Environ. Urban Syst.*, vol. 55, pp. 44–54, Jan. 2016.
[29] D. Chen, C. S. Ong, and L. Xie, "Learning points and routes to recommend trajectories," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2016, pp. 2227–2232.
[30] G. Cui, J. Luo, and X. Wang, "Personalized travel route recommendation using collaborative filtering based on GPS trajectories," *Int. J. Digit. Earth*, vol. 11, no. 3, pp. 284–307, Mar. 2018.
[31] S. Jiang, X. Qian, T. Mei, and Y. Fu, "Personalized travel sequence recommendation on multi-source big social media," *IEEE Trans. Big Data*, vol. 2, no. 1, pp. 43–56, Mar. 2016.
[32] P. Bolzoni, S. Helmer, K. Wellenzohn, J. Gamper, and P. Andritsos, "Efficient itinerary planning with category constraints," in *Proc. 22nd Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL)*, 2014, pp. 203–212.
[33] G. Cai, K. Lee, and I. Lee, "Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos," *Expert Syst. Appl.*, vol. 94, pp. 32–40, Mar. 2018.
[34] H.-T. Chang, Y.-M. Chang, and M.-T. Tsai, "ATIPS: Automatic travel itinerary planning system for domestic areas," *Comput. Intell. Neurosci.*, vol. 2016, Dec. 2016, Art. no. 1281379.
[35] S. Rani, K. N. Kholidah, and S. N. Huda, "A development of travel itinerary planning application using traveling salesman problem and K-Means clustering approach," in *Proc. 7th Int. Conf. Softw. Comput. Appl. (ICSCA)*, 2018, pp. 327–331.
[36] S. Mancini and G. Stecca, "A large neighborhood search based matheuristic for the tourist cruises itinerary planning," *Comput. Ind. Eng.*, vol. 122, pp. 140–148, Aug. 2018.
[37] J. Du, L. Li, and X. Li, "Data-driven travel itinerary with branch and bound algorithm," in *Proc. DASC/PiCom/DataCom/CyberSciTech*, Aug. 2018, pp. 1046–1053.
[38] M. Figueredo, J. Ribeiro, N. Cacho, A. Thome, A. Cacho, F. Lopes, and V. Araujo, "From photos to travel itinerary: A tourism recommender system for smart tourism destination," in *Proc. IEEE 4th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2018, pp. 85–92.
[39] D. Yang, T. Lu, H. Xia, B. Shao, and N. Gu, "Making itinerary planning collaborative: An AST-based approach," in *Proc. IEEE 20th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2016, pp. 257–262.
[40] Y.-L. Hsueh and H.-M. Huang, "Personalized itinerary recommendation with time constraints using GPS datasets," *Knowl. Inf. Syst.*, vol. 60, no. 1, pp. 523–544, Jul. 2019.
[41] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "YFCC100M: The new data in multimedia research," *Commun. ACM*, vol. 59, no. 2, pp. 64–73, Jan. 2016.
[42] Wikipedia. *List of Sights and Historic Places in Budapest*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_sights_and_historic_places_in_Budapest
[43] Wikipedia. *Category: Tourist Attractions in Edinburgh*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/CategoryTourist_attractions_in_Edinburgh
[44] Wikipedia. *List of Tourist Attractions in Toronto*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_tourist_attractions_in_Toronto
[45] Wikipedia. *Category: Tourist Attractions in Glasgow*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/CategoryTourist_attractions_in_Glasgow
[46] Wikipedia. *Tourism in Perth*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Tourism_in_Perth
[47] Wikipedia. *Category: Tourist Attractions in Osaka*. Accessed: Aug. 12, 2019. [Online]. Available: https://en.wikipedia.org/wiki/CategoryTourist_attractions_in_Osaka

PHATPICHA YOCHUM received the B.S. degree in software engineering from Mae Fah Luang University, Thailand, in 2009, and the M.S. degree in information technology from Rangsit University, Thailand, in 2017. She is currently pursuing the Ph.D. degree in information and communication engineering with the Guilin University of Electronic Technology, China. Her research interests include knowledge graphs, network embedding, and recommendation systems.

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

TIANLONG GU received the Ph.D. degree from Zhejiang University, China, in 1996. From 1998 to 2002, he was a Research Fellow at the School of Electrical and Computer Engineering, Guilin University of Technology, Australia, and Postdoctoral Fellow with the School of Engineering, Murdoch University, Australia. He is currently a Professor with the School of Computer Science and Information Security, Guilin University of Electronic Technology, China. His research interests include formal methods, data and knowledge engineering, and software engineering.



MANLI ZHU received the bachelor's degree in software engineering from Zhoukou Normal University, in 2013, and the M.S. degree in computer science and technology from the Guilin University of Electronic Technology, China. Her research interests include knowledge graphs, knowledge graph representation, and recommendation systems.