Physica A 525 (2019) 798–808

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A movement model for air passengers based on trip purpose

Feihu Huang [a,b], Shaojie Qiao [a,*], Jian Peng [b,*], Bing Guo [b], Xi Xiong [c], Nan Han [d]

[a] School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China
[b] College of Computer Science, Sichuan University, Chengdu 610065, China
[c] School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China
[d] School of Management, Chengdu University of Information Technology, Chengdu 610103, China

HIGHLIGHTS

- Movement model for air passengers based on trip purpose and memory effect.
- Influences of trip purpose calculated by advance booking time and departure date type.
- Gross domestic product, number of travelers, individual factors used to compute utility of city attraction.

ARTICLE INFO

Article history:
Received 24 October 2018
Received in revised form 28 January 2019
Available online 6 April 2019

Keywords:
Movement model
Human mobility
Trip purpose
Memory effect
Throughput prediction

ABSTRACT

The study of human mobility has contribution to understanding the law of human spatial movement and has obtained a lot of attention in recent years because of its applications in location-based social networks (LBSNs). The existing works mainly focus on analyzing the human movement on the ground due to the limitation of acquiring data. In addition, existing models ignore the influence of individuals' trip purpose on human movement. In this study, a novel movement model for air passengers based on trip purpose is proposed. There are two essential factors in the proposed model: (1) Trip purpose. We firstly verify the effects of trip purpose by analyzing air passengers' reservation records and then assume that the travel of air passengers is driven either by business purpose or tourism purpose. Based on the analysis of trip purpose, a method based on expected utility theory is proposed when calculating the probability of choosing airports for passengers. (2) Memory effect. In order to depict the movement characteristics of passengers, we take into full consideration the memory effect when determining whether or not to choose the candidate airports from historical airports or new airports which have not be visited before. Extensive experimental results demonstrate that the proposed movement model has obtained good performance on the individual as well as the population levels, and outperforms the state-of-the-art Gravity model and Radiation model.

1. Introduction

In recent years, the study of human mobility has received intensive attention. The access of large collection of human movement data strongly contributes to the research of human mobility [1–3]. Human mobility not only helps us understand the laws of human movement, but also is an important issue for the research related to the human

* Corresponding authors.
E-mail addresses: sjqiao@cuit.edu.cn (S. Qiao), jianpeng@scu.edu.cn (J. Peng).

https://doi.org/10.1016/j.physa.2019.04.040

movement [4,5], such as Point-of-Interest recommendation which is an active topic in location-based social networks, urban planning [6–8], traffic forecasting [2,9,10] and epidemic control [11–13]

Recently, some researchers have proposed some human mobility prediction models. The Gravity model [14] which integrates the distance between two locations and the number of commuters of locations based on the Newton's law of gravity, is one of the famous models. Noulas et al. [15] extended the Gravity model to handle the problem of human movement in urban areas by calculating the number of places between the departure place and destination. Another popular model is the radiation model [16], which is a non-parameter model and has several improved models. For instance, Simini et al. [17] proposed a continuum approach for human mobility by combining the radiation model and the intervening opportunity model. In order to deal with the problem of estimating the commuting flows at different spatial scales without using the calibration data, Yang et al. [18] employed a scaling parameter α to calculate the commuting flows.

However, the existing studies of human mobility mainly focus on the human activities on the ground and barely designing the models for air passengers. For example, the trajectories of currency circulation are used to investigate the scaling laws of human traveling [19]. The mobile phone data is also used to explore the mobility patterns of individuals [6,20]. In addition, the trajectory data of taxis [21], the smart card records of public transports [22,23] and GPS positioning data [24] has been extensively used to analyze the human movement. The studies on aviation data mainly focus on throughput forecasting [25] and analyzing the essential factors on throughput [26] and aviation networks [27]. In terms of the mobility of air passengers, to the best of our knowledge, only our previous work [28] analyzed the characteristics of crowd movement of air passengers.

Motivation. Most of the previous models mainly focus on studying the characteristics of human movement (such as the geographical proximity [29,30], the memory effect [31,32], exploration and preferential return [4,6]), but the effects of trip purpose are ignored. The trip purpose is a very important factor on human mobility. For instance, one goes to a place either for work (meeting customers or business meetings) or for routines (visiting friends or shopping). One will choose a place to achieve her/his purpose. That is to say, human mobility is driven by her or his purpose [33].

Contribution. In this study, we aim to investigate the mobility of air passengers based on air passengers' reservation records. We should take into consideration two questions: (1) Do air passengers' travel have relationship with their trip purpose? (2) How to model the movement characteristics of air passengers? The original contributions are given as follows:

(1) We propose a novel movement model for air passengers. The model takes into account trip purpose as well as memory effect. We give the theoretical analysis of each metric and conduct experiments from individual level and population level to verify the performance of the proposed model.

(2) We verify the influences of trip purpose on passengers' travel by calculating the relationship between the time of advance booking and the type of departure date.

(3) We propose a method to calculate the choosing probability for each candidate airport. In terms of the method, the urban economic factors including GDP (Gross Domestic Product) and the number of travelers, and individual factors (trip purposes) are used to calculate the utility which corresponds to the attraction of cities.

The rest of this paper is organized as follows. The characteristics of movement of air passengers is presented in Section 2. The detail of our model is introduced in Section 3. We make the theoretical analysis of the proposed model in Section 4. The experimental studies are discussed in Section 5. Finally, we discuss this work in Section 6.

2. Characteristics of movement

2.1. Time interval

The time interval is an important property of human movement. The time difference between two adjacent trips can reflect passenger's travel frequency.

Fig. 1 illustrates the distribution of time interval (unit is day) of air passengers. As we can see, the distribution agrees with the truncated power law of which *pdf*(probability density function) is defined as $f(x) \sim x^{-\alpha}e^{\lambda x}$ with $\alpha = 1.08$ and $\lambda = 0.08$. We use the Python package in [34] to obtain the parameters of fitting function. The minimum time interval is 0, which means some passengers (2.49%) either return to their departure place or start a new trip in one day. This result agrees with the characteristics of air passengers. For example, if one goes to "Beijing" from "Chengdu" to attend a business meeting, she/he may choose to return to "Chengdu" after meeting or fly to another city for official business.

2.2. Trip distance

In addition, we calculate the distance distribution of passengers' trips (as shown in Fig. 2(a)) and the distance distribution of the route network (as shown in Fig. 2(b)), respectively. The unit of distance is kilometer.

We can see that the distance distributions obey the negative binomial distribution (defined as $f(x) = C_{x+r-1}^{r-1}p^r(1-p)^x$). The result agrees with the study given in [22] and [23]. In [23] the researchers studied the mobility of passengers in London subway. In [22], the authors explored the mobility patterns of public transport passengers. According to the conclusions

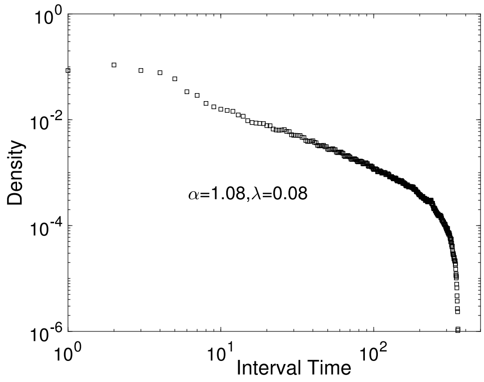
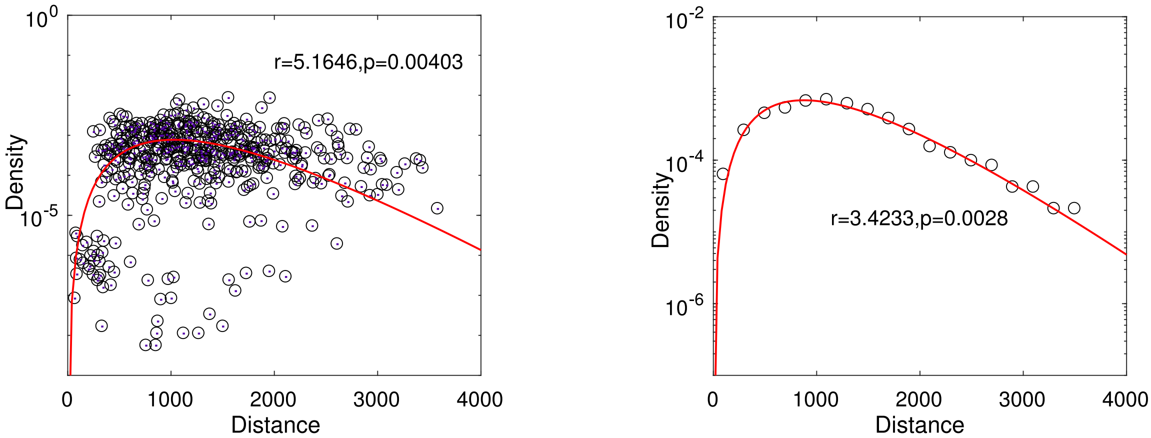


Fig. 1. Distribution of time interval.



(a) Trip records (b) Route network

Fig. 2. Distance distribution.

obtained from [22] and [23], we infer that the travel based on transportation systems can be depicted uniformly by the negative binomial distribution. For air passengers, the same distribution of trip distance and route distance implies that the structure of a route network can affect the movement of passengers. The values of corresponding parameters between the two distance distributions are different. In Fig. 2(a), the values of parameters are $r = 5.1646$ and $p = 0.00403$. In Fig. 2(b), the values of parameters are $r = 3.4233$ and $p = 0.0028$. This indicates that the structure of route network is not a single influence factor for air passengers' movement [22,23].

2.3. Trip purpose

In order to explore whether air passenger's travel is greatly affected by trip purpose, we illustrate passengers' distribution of days (which is the time between booking ticket and departure) in Fig. 3, and the relationship between the days of advance ticket booking and the type of departure date in Fig. 4. The distribution of days obeys the stretched exponential (defined by $p(x) = e^{-\lambda x^{\beta}}$) with $\lambda = 0.2$ and $\beta = 0.64$. We know the price of ticket is related to the time of ticket booking. In general, the longer the booking time is before traveling, the bigger the discount is. However, most air passengers have a short time interval between ticket booking and their trip, as shown in Fig. 4. These passengers might not care about the price of tickets because they are on a business trip. Fig. 4 shows that the trips of most passengers occur on weekdays and the percentage of trips which occur on holidays becomes more frequent when the days of advance booking increases. We can infer that the passengers who has a long time interval between booking time and travel time are travelers for tourism purpose. This can be explained by the reason that: generally speaking, people will make a plan in advance before starting a travel and they care about the ticket price. Therefore, with the increase of days of advance ticket booking, the number of passengers who choose travel on holidays grows. In summary, the trip purpose of air passengers can be verified by the time interval between ticket booking and their trip.

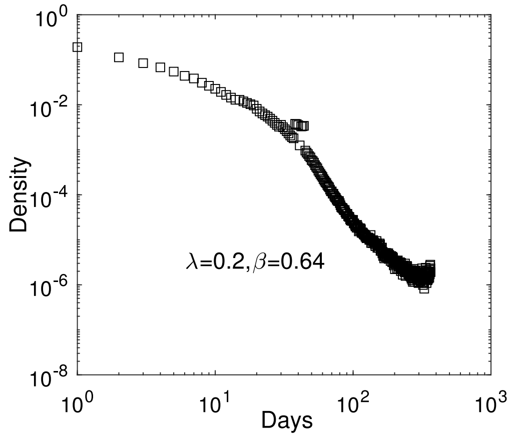


Fig. 3. Distribution of days.

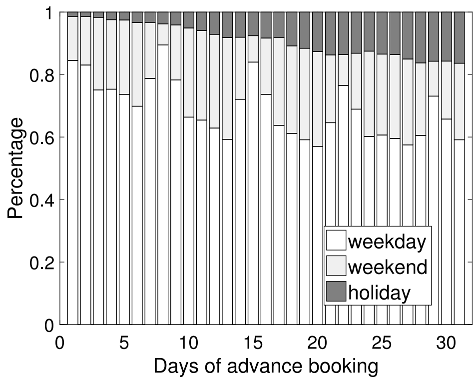


Fig. 4. Relationship between the time of advance booking and different kinds of departure date.

3. Model introduction

As discussed in Section 2, trip purpose will influence the choosing of people's destination [35,36]. It is worthwhile to note that the purposes of activities are different. Due to the incompleteness of data, the trip records of air passengers cannot provide the purpose of each trip. However, according to the aforementioned analysis of the days of advance ticket booking, we can approximately classify the trip purpose of air passengers into two categories: business purpose and tourism purpose.

When choosing a place, we wish the destination could satisfy our demand. In this study, the capability of a place satisfying people's demand is called "Utility". Specially, from the perspective of economics, air passengers hope that they can obtain the maximum utility from the destination. According to the expected utility theory [37], U_{ij} is defined as the utility user i obtained from place j, which is calculated by the following equation:

$$U_{ij} = p \cdot u(x_{j1}) + (1 - p) \cdot u(x_{j2}) \tag{1}$$

where p is the probability that one selects city j for business purpose. For air passengers, the trip purposes are business or tourism, then $1 - p$ is the probability for tourism purpose. Note that, the two probabilities p and 1-p correspond to the attraction of business and tourism in a city, respectively. Therefore, we use the *Beta* distribution to calculate the choosing probability. Then, Eq. (1) can be transformed into the following equation:

$$U_{ij} = Beta(\alpha_1, \alpha_2) \cdot u(x_{j1}) + (1 - Beta(\alpha_1, \alpha_2)) \cdot u(x_{j2}) \tag{2}$$

In Eq. (2), the parameters α_1 and α_2 correspond to the GDP and the number of travelers of the city, respectively. $u(\cdot)$ is the utility function. In this paper, the utility function is $u(x) = \sqrt{x}$. x_{j1} corresponds to GDP and x_{j2} corresponds to the number of travelers. However, the values of GDP and number of travelers have different meaning. For example, one city's GDP is 100 billion RMB, and there are 10 million people come to this city for travel purpose. Obviously, the value of GDP is much greater than the number of travelers. So we need to regularize these two parameters. Specially, the values of α_1 and α_2 after regularization will be between 0 and 1. According to the properties of probability distribution function of *Beta* distribution, if α_1 and α_2 are both less than 1, the curve of probability distribution function follows a U-shaped pattern, we get a bimodal distribution with spikes at 0 and 1. If α_1 and α_2 are both greater than 1, the distribution is unimodal. Therefore, we need a parameter to ensure that the values of α_1 and α_2 after regularization are greater than 1. In this study, we regularize the GDP and traveler number of city by the following equation:

$$\alpha_j = a \cdot \frac{x_j}{max(X)} \tag{3}$$

where a is the parameter which ensures the values of α and β after regularization are greater than 1. The way to set the value of a is based on the minimum value of the normalization parameter. According to data sets in this paper, we set a to 10^4.

For each city, its adjacent cities (denoted by V) are the cities that are navigable with it. Therefore, the probability that a passenger travels from city i to j is calculated by the following equation:

$$p_{ij} = \frac{U_{ij}}{\sum_{j\in V} U_{ij}} \tag{4}$$

Algorithm 1 The simulation procedure of the proposed model

Input:
λ: memory effect;
N: the number of total airports;
L_h: the sequence of passenger's visited airports.
Output:
L_s: the set of airports generated by simulations.
1: Calculate T the number of trips w.r.t. a passenger by L_h;
2: Calculate H the home airport in L_h;
3: Set the departure airport i as H and the current number of trips t to 1;
4: Add i to the sequence L_s;
5: **while** $t <= T$ **do**
6: **if** $t == 1$ **then**
7: p_{new}=1;
8: **else**
9: Calculate p_{new} by Eq. (5);
10: **end if**
11: **if** $rand < p_{new}$ **then**
12: Calculate the candidate airports C from i's adjacent airports that the passenger has not visited; //exploring new city
13: **else**
14: Calculate the candidate airports C from historical airports that are navigable with i; //visiting historical city;
15: **end if**
16: Choosing the next visiting airport j by Eq. (2) on the basis of C;
17: Add j to the sequence L_s;
18: Update the next departure airport i=j;
19: Update the number of trips t=$t+1$;
20: **end while**

Memory effect is another important factor we should take into consideration. Specially, for air passengers, they have a high probability of visiting the cites which they had visited before, which agrees with the behaviors of air passengers. The aforementioned analysis indicates that most people travel by plane for business purpose. Therefore, going to historical cities to meet clients or cooperative partners is a common business behavior. We define p_{old} as the probability of visiting historical cities. The probability of exploring new city p_{new} is defined below:

$$p_{new} = \frac{N - S(t)}{N - S(t) + t + \lambda} \tag{5}$$

where N represents the number of total airports, $S(t)$ is defined as the cumulative number of cities visited after t trips. t is the number of trips, and λ is used to reflect the influences of memory effect.

Algorithm 1 describes the simulation procedure of the proposed model. It can be divided into three phases: (1) Initialization of parameters (Steps 1–4), in which calculating the number of trips w.r.t. a passenger takes T operations and finding home airport needs T operations. (2) Choosing the candidate airports (Steps 6–15). The worst situation is choosing the candidate airports from historical airports if the current airport is navigable with all the airport. It takes $|L_s| \cdot N$ operations. (3) Choosing the next visiting airport on the basis of visiting probability (Step 16), which needs $|C|$ operations. Therefore, the time complexity of Algorithm 1 is $O(K(2 \cdot T^2 + T \cdot |L_s| \cdot N + T \cdot |C|))$. Since T, N, $|C|$ and $|L_s|$ are constant with small values, Algorithm 1 can be viewed as linear in $O(K)$, where K is the number of passengers.

It is worthwhile to note that the home airport of a passenger in the proposed model is the most frequent airport appeared in the passenger's travel records. In the data sets, passenger's travel trajectories are discontinuous. For example, the previous travel record of a passenger is from city A to B, and the current record is from city A to C. And, the record that the passenger returns to city A from city B is missing. The reason is that passenger can choose different airline companies for traveling. Therefore, the missing travel record that the passenger returns to city A from city B may appear in other airline's records. In the simulation procedure, we view passenger's travel trajectory as a continuous procedure, which implies the next departure city should be city B if the current travel record is from city A to city B.

4. Theoretical analysis

In order to verify the performance of our proposed model, we will formalize the statistical behaviors of air passengers at the individual as well as the population levels. At the individual level, we seek to discuss: (1) the cumulative number of cities visited for t times and (2) time interval distribution of visiting historical city. At the population level, we focus on the following quantities: (1) the total arrival throughput of airports and (2) the throughput of the routes. In this section, we will analyze the fundamental theories of the proposed model. The experimental results are given in the following section.

4.1. Individual level

Based on the aforementioned discussion, the probability of exploring a new city is defined by the following equation:

$$p_{new} = \frac{dS(t)}{dt} = \frac{N - S(t)}{N - S(t) + t + \lambda} \tag{6}$$

Given the initial condition $S(1) = 1$, we can obtain the dependence of $S(t)$ on the number of trips t as below:

$$S(t) = N + t + \lambda - \sqrt{(N+\lambda)^2 + (\lambda + t)^2 - (\lambda + 1)^2} \tag{7}$$

For a given city i, f_i is the frequency of visiting i, the probability of i visited by passenger at the tth trip is defined as follows:

$$p_i = \frac{t + \lambda}{N - S(t) + t + \lambda} \cdot \frac{f_i}{t} \tag{8}$$

Song et al. [6] defined f_i is proportional to t/r_i, where r_i denotes the time step when the location i was first visited. Applying Eq. (7) and $f_i \propto t/r_i$ into Eq. (8), We can obtain the following equation:

$$p_i \propto \frac{t+\lambda}{\sqrt{(N+\lambda)^2 + (\lambda+t)^2 - (\lambda+1)^2}} \cdot \frac{1}{t} \cdot \frac{t}{r_i} = \frac{t+\lambda}{\sqrt{(N+\lambda)^2 + (\lambda+t)^2 - (\lambda+1)^2}} \cdot \frac{1}{r_i} = \frac{t+\lambda}{t+\lambda+C^*} \cdot \frac{1}{r_i} \propto \frac{1}{r_i} \tag{9}$$

where C^* is a small constant. We ignore the effect of C^* to make an approximation. According to p_i, the return time distribution to any location P_τ is defined below:

$$P_i(\tau) = p_i(1 - p_i)^{\tau - 1} \tag{10}$$

where τ is the interval frequency. For example, a passenger's visiting sequence to city A is {2, 5, 6, 9, 14}. The element in the set means the user visited city A at the ith trip. Therefore, the time frequency of the return time distribution is {3, 1, 3, 6}. It is worthy to note that we calculate the time difference (unit is day) between two adjacent trips in Fig. 1, in which the time interval is different from the interval frequency in Eq. (10). For all the visited cities, the return time distribution is described by the following equation:

$$P(\tau) = \frac{1}{f(\tau)} \int_1^{S(t)} \int_1^{T-\tau} \frac{1}{r_i} \left(1 - \frac{1}{r_i}\right)^{\tau-1} dr_i ds \tag{11}$$

The above equation is a function agreeing with power law. $P(\tau)$ is calculated by double integral operation since $P_i(\tau)$ is relevant with the number of trips t and the number of airports s. The r_i is less than $T - \tau$, i.e., $1 < r_i < T - \tau$, where T is the maximum number of trips. The integral range for s is between $[1, S(t)]$. Note that $S(t)$ is larger than the number of airports that are visited after τ times at the tth trip.

Furthermore, with the increase of τ, the number of the airports visited by passengers after τ times decreases. So the integral for s from 1 to $S(t)$ leads to a larger value of $P(\tau)$ than the actual value. In order to cope with the problem, we use $f(\tau) = \frac{1}{\tau}$ to estimate the relationship between the actual number of airports (which visited by passengers after τ times) and $S(t)$. Experiments indicate that the function $f(\tau)$ in Eq. (11) can reduce the error of $P(\tau)$ obtained by integral operation for s from 1 to $S(t)$.

Table 1
Description of the datasets.

Passengers	15,616,181
Airports	103
Routes	903
Time span	1/8/2014 to 12/30/2014
Attributes	Departure airport, destination airport, flight number, departure time, booking time, seat number
Record	SWA,CTU,CA4374,12/25/2014,11/11/2014,21K

4.2. Population level

According to Eq. (2), the probability that a passenger chooses city j for business purpose is p. At the popular level, the expected probability $E(p)$ is defined as follows:

$$E(p) = \int_0^1 p \cdot \frac{p^{\alpha_1 - 1}(1-\alpha_1)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} dp = \frac{\alpha_1}{\alpha_1 + \alpha_2} \tag{12}$$

Therefore, $E(U_{ij})$ (the expected utility for passenger from city i to j) is defined by the following equation:

$$E(U_{ij}) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \cdot u(x_{j1}) + \frac{\alpha_2}{\alpha_1 + \alpha_2} \cdot u(x_{j2}) \tag{13}$$

Based on the GDP and the number of travelers, a passenger visits city j from city i with the following probability:

$$\overline{p}_{ij} = \frac{E(U_{ij})}{\sum_{j \in V} E(U_{ij})} \tag{14}$$

For city i, we assume that its airport throughput is G_i, in which there are G_{ij} passengers depart from city i to city j. The process of G_{ij} passengers departing from city i to city j can be viewed as the Bernoulli process [16]:

$$P(G_{ij}) = \frac{G_i!}{G_{ij}!(G_i - G_{ij})!} \overline{p}_{ij}^{G_{ij}} (1 - \overline{p}_{ij})^{G_i - G_{ij}} \tag{15}$$

Eq. (15) is a binomial distribution with the following average route throughput:

$$\overline{G}_{ij} = \overline{G}_{ij} = G_i \cdot \overline{p}_{ij} \tag{16}$$

where $\overline{G}_{ij}$ is the average route throughput from city i to j. Based on $\overline{G}_{ij}$, the total number of passengers that arrive at city j can be calculated by the following equation:

$$\overline{G}_j = \sum_i \overline{G}_{ij} \tag{17}$$

At the population level, we view $\overline{G}_j$ and $\overline{G}_{ij}$ as the theoretical results of the total arrival throughput of airport and the throughput of route, respectively.

5. Experiments

5.1. Data description

The data used in this study is obtained from the airline of China. The data covers the flight routes of the year of 2014, which contains the attributes of passenger's flight number, seat number, departure airport, departure time, destination airport and booking time. After eliminating the incorrect data and temporary route, we obtain 42,543,756 travel records of 15,616,181 passengers, a total number of 103 airports and 903 routes. Table 1 shows the details of the data. The values of GDP and the number of travelers are obtained from the portal of http://www.tjcn.org/tjgb/.

5.2. Similarity metric

SSI can identify the similarity between two samples. In human movement models, it is used to evaluate the prediction performance by calculating the similarity between empirical data and prediction value [29,38]. SSI is defined as follows:

$$SSI = \frac{1}{W^2} \sum_i^W \sum_j^W \frac{2min(\overline{G}_{ij}, G_{ij})}{\overline{G}_{ij} + G_{ij}} \tag{18}$$

where $\overline{G}_{ij}$ is the throughput predicted by different models, G_{ij} is the real throughput from empirical data, and i and j refer to the corresponding cities. W is the number of samples. SSI $= 1$ indicates that all the $\overline{G}_{ij}$ are equal to G_{ij}, otherwise SSI is close to 0.

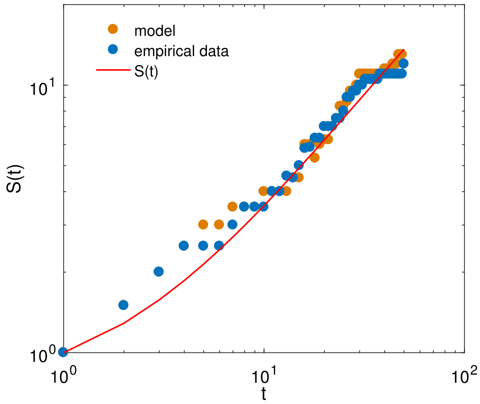


Fig. 5. Cumulative number of cities visited by t times.

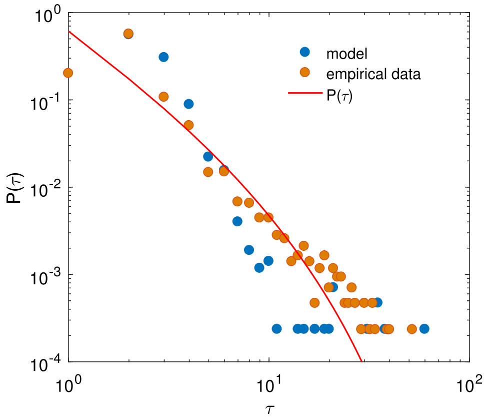


Fig. 6. Return time distribution of τ.

5.3. Baseline model

In the experiment, we have two baseline models, the radiation model and the gravity model. The probability of passenger from city i to j in the radiation model is defined as follows [16]:

$$p_{ij} = \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \tag{19}$$

where m_i is the population of the city where the airport i is located; n_j is the population of the city where the airport j is located; s_{ij} is the sum of the population of the airports which have routes to airport i.

The probability of passenger from city i to j in the gravity model is defined as follows [14]:

$$p_{ij} = \frac{m_i n_j}{d(i,j)^b} \tag{20}$$

Using the same method in [18], the distance function is a power function, we have b $= -0.13$. m_i and n_j have the same meaning as defined in Eq. (19). $d(i,j)$ is the distance between the city where the airport i is located at and the city where the airport j is located at.

5.4. Results

For the experiments at individual level, we sample 10^3 passengers from the real data and then begin the simulation procedure for each passenger via Algorithm 1. In the simulation procedure, λ is set to 250. For each passenger, the number of trips is the actual number which is calculated from her/his travel trajectories. For the experiments at population level, for any airport i, the number of passengers that depart from i is its throughput which can be obtained from the empirical data. And then $\overline{G}_{ij}$ and $\overline{G}_j$ are calculated by Eqs. (16) and (17), respectively.

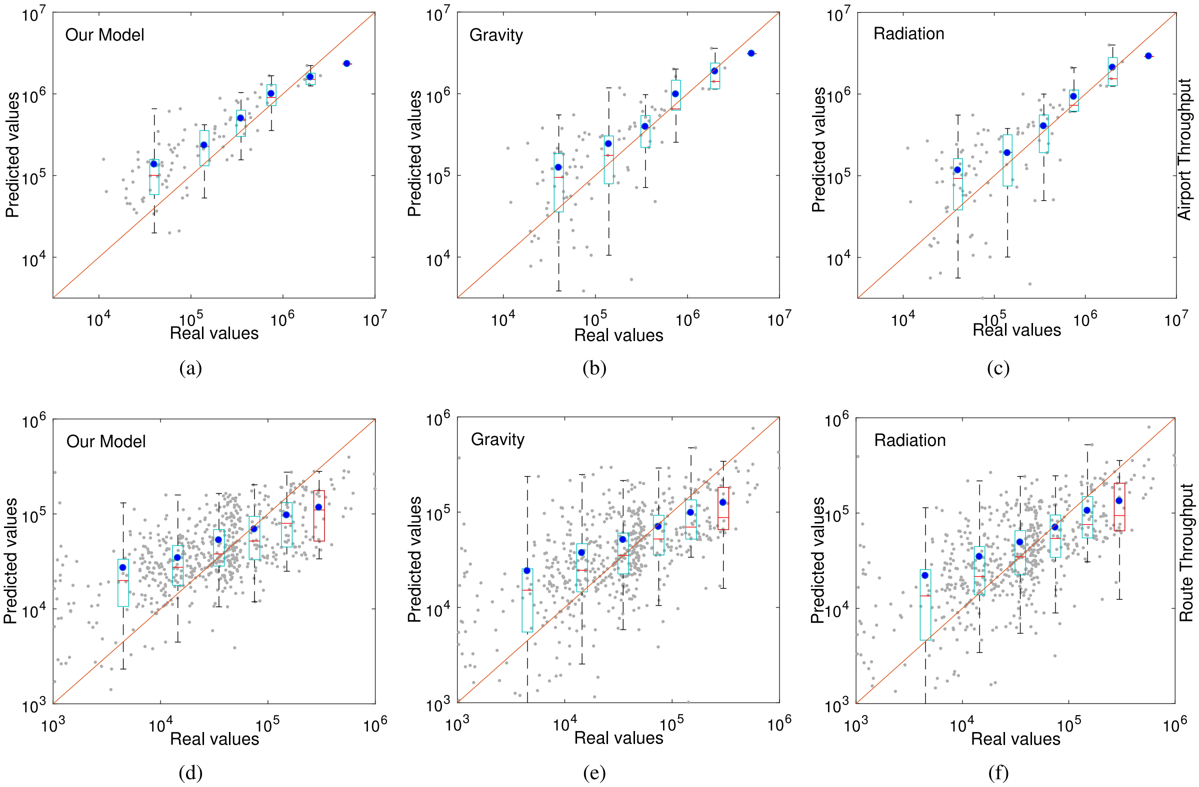


Fig. 7. Throughput comparison of different models.

At the individual level, Fig. 5 shows the relationship between the cumulative number of cities visited by passengers and t. The proposed model obtains the algebraic increase of $S(t)$ as t increases (orange dots). The results excellently agree with the empirical result (blue dots) and theoretical analysis (red line). The algebraic increase of $S(t)$ also obeys the behaviors of passengers: the more chances to travel, the more places to visit. It is worth to note that the growth rate of $S(t)$ is less than that of t. This is because of the influence of memory effect [32]. In our model, we set a larger value for the memory effect $\lambda = 250$. According to Eq. (5), the larger memory effect will lead to a less probability of exploring new places, i.e., a greater probability to visit historical places. Unlike the behaviors of random walks, air passengers have a larger probability of returning to departure place after traveling.

Fig. 6 illustrates the return time distribution of τ. The proposed model also obtains an agreement with the empirical result (blue dots), together with the theoretical analysis (red line). $P(\tau)$ has an algebraic decay when the interval frequency τ increases. In addition, $P(\tau)$ obeys the power law, which means the proportion of larger τ (tail) is very small. That is to say, passengers usually revisited the places that they have visited before with small interval frequency. The algebraic decay of $P(\tau)$ is due to the memory effect. The number of trips is an another important factor. For most of air passengers, the number of trips during one year are not large, so the value of $P(\tau)$ with a large value of τ is small.

At the population level, Figs. 7(a)–7(c) show the results of annual airport throughput, and Figs. 7(d)–7(f) illustrate the results of annual route throughput. The y-axis corresponds to the data generated by our proposed model. The x-axis corresponds to the empirical data. The route throughput is the total throughput of the flights from the departure to the destination city. The airport throughput is the number of passengers arriving at this airport. In Fig. 7, the better performance of prediction corresponds to the phenomenon that the scatter points are close to the line $y = x$. A box is marked in light green if the line $y = x$ lies between 10% and 90%, and in red otherwise. According to Fig. 7, the performance of the airport throughput of these three models are better than those of the route throughput. The route throughput predicted by simulation is less than the empirical data. All these three models have the same drawback. This is because of the variety of influences on the throughput [26]. In summary, we can conclude that the predicted airport and route throughput produced by the proposed model are approximate to empirical data that is satisfying.

In order to estimate the prediction performance of the route throughput among the proposed model, Gravity model and Radiation model, the SSI metric [39] is used in this study. Table 2 shows the values of SSI for different models. We calculate the SSI value of annual throughput for each airport and route, and of the daily throughput for each airport and route, respectively. The results imply that the proposed model outperforms the Gravity model and Radiation model on both of the airport throughput and route throughput.

Table 2
SSI for different models.

	Annual		Daily	
	Airport	Route	Airport	Route
Model	**0.6262**	**0.6442**	**0.3054**	**0.3675**
Gravity	0.6184	0.6438	0.233	0.3627
Radiation	0.611	0.6418	0.234	0.3624

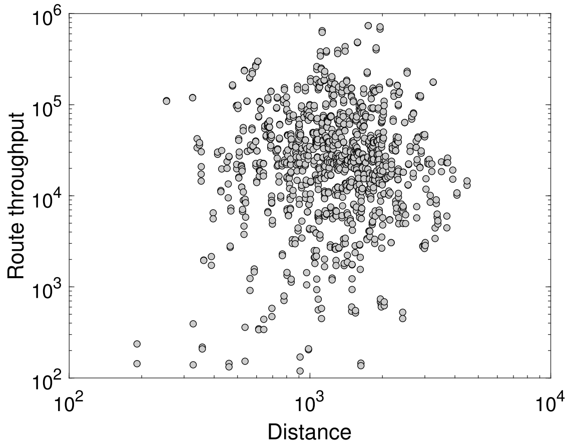


Fig. 8. Relationship between route throughput and distance.

6. Discussions

There are several factors affecting the human movement. Population density is an important factor for human movement since the number of opportunities at a location is proportional to its population [16,29]. Therefore, the place having high population density implies more employment opportunities and can attracts several people to work there. The distance is another essential factor [40]. Specially, for the trips inside the city, travelers will take into account the time cost between two locations [28]. In addition, social connections also have effects on human movement. Existing studies indicate that friends indeed are much closer in behavior patterns than strangers [41,42]. That is to say, one prefers to travel the places which visited by their friends. In recent years, with the accumulation of large number of human movement data, researchers indicate that the temporal characteristics of human mobility agree with the power law [43]. Memory effect is an useful explanation for this phenomenon. Some human movement models treat memory effects as an essential ingredient [31,32].

To the best of our knowledge, travel by air becomes popular and the population flow between cites can be influenced by several factors. In our study, we classify the factors into two categories: urban economic factors (such as GDP, population and the tertiary industry output) and individual factors (such as business and tourism). Therefore, we adopted the expected utility theory to estimate the city attraction for passengers. The proposed model integrates the urban economic factor and individual factor. Based on the statistical analysis of empirical data, we assume that business and travel are the main trip purposes of air passengers. GDP and the number of travelers correspond to the metric of business utility and tourism utility. Different from the movement on the ground, passengers travel by air are because of the long distance between two cities. As shown in Fig. 8, the route throughput does not degrade with the increase of route length. It is worth to note that although the distance has no effect on travel, the travel of air passengers relies on the route network. Furthermore, memory effect is taken into consideration in the proposed model since it is one typical characteristic of human behavior.

The proposed model is verified from two aspects. At individual level, human mobility consists of spatial characteristic and temporal characteristic [28,43]. So we investigate the return time distribution of visiting historical city for temporal characteristic. In the simulation, we calculate the interval frequency to investigate the temporal characteristic as the same did in [32]. Due to the memory effect, exploring new place is considered as the evaluation method of individual travel. At population level, the aggregative mobility of air passengers can mainly be reflected by the throughput. Therefore, we calculate the SSI values of airport throughput and route throughput to evaluate the performance of the proposed model. In summary, the proposed movement model can well describe the movement characteristics of air passengers at individual level and population level. We believe the methods proposed in this research work to contribute to a larger effort targeted at advancing the study of human mobility prediction.

Acknowledgments

This work is partially supported by the State Key Program of National Natural Science Foundation of China under Grant No. 61332001; the National Natural Science Foundation of China under Grant Nos. 61772091, 61802035, 61772352, 61472050; the Sichuan Science and Technology Program under Grant Nos. 2018JY0448, 2019YFG0106, 2019YFS0067, 2019ZDZX0045, 2019ZDZX0010, 2018ZDZX0010, 2017GZDZX0003, 2018JY0182; the Innovative Research Team Construction Plan in Universities of Sichuan Province under Grant No. 18TD0027; the Natural Science Foundation of Guangxi under Grant No. 2018GXNSFDA138005; the Scientific Research Foundation for Advanced Talents of Chengdu University of Information Technology under Grant Nos. KYTZ201715, KYTZ201750; the Scientific Research Foundation for Young Academic Leaders of Chengdu University of Information Technology under Grant No. J201701; the Soft Science Foundation of Chengdu under Grant No. 2017-RK00-00053-ZF.

References

[1] C. Fan, Y. Liu, J. Huang, Z. Rong, T. Zhou, Correlation between social proximity and mobility similarity, Sci. Rep. 7 (1) (2017) 11975.
[2] Z.D. Zhao, Z.G. Huang, L. Huang, H. Liu, Y.C. Lai, Scaling and correlation of human movements in cyberspace and physical space, Phys. Rev. E 90 (5) (2014) 050802–050802.
[3] S. Qiao, N. Han, J. Wang, R.H. Li, L.A. Gutierrez, X. Wu, Predicting long-term trajectories of connected vehicles via the prefix-projection technique, IEEE Trans. Intell. Transp. Syst. 2018 (7) (2018) 2305–2315.
[4] J. Peng, M. You, F. Huang, B. Tong, Civil aviation passenger mobility model based on the potential trip purpose, Adv. Eng. Sci. 49 (6) (2017) 107–113.
[5] F. Huang, X. Xiong, J. Peng, B. Guo, B. Tong, Rca: A route city attraction model for air passengers, Physica A 491 (2018) 887–897.
[6] C. Song, T. Koren, P. Wang, A.L. Barabási, Modelling the scaling properties of human mobility, Nat. Phys. 6 (10) (2010) 818–823.
[7] S. Qiao, N. Han, W. Zhu, L.A. Gutierrez, Traplan: an effective three-in-one trajectory-prediction model in transportation networks, IEEE Trans. Intell. Transp. Syst. 16 (3) (2015) 1188–1198.
[8] S. Qiao, N. Han, J. Zhou, R.H. Li, C. Jin, L.A. Gutierrez, Socialmix: A familiarity-based and preference-aware location suggestion approach, Eng. Appl. Artif. Intell. 68 (2018) 192–204.
[9] Y. Ren, M. Ercsey-Ravasz, P. Wang, M.C. González, Z. Toroczkai, Predicting commuter flows in spatial networks using a radiation model based on temporal ranges, Nature Commun. 5 (2014) 5347.
[10] S. Qiao, D. Shen, X. Wang, N. Han, W. Zhu, A self-adaptive parameter selection trajectory prediction approach via hidden markov models, IEEE Trans. Intell. Transp. Syst. 16 (1) (2015) 284–296.
[11] F. Xiong, Z.-Y. Li, Effective methods of restraining diffusion in terms of epidemic dynamics, Sci. Rep. 7 (1) (2017) 6013.
[12] D. Wang, S.-J. Cheng, A two-stage broadcast message propagation model in social networks, Physica A 462 (2016) 1286–1293.
[13] S. Qiao, N. Han, Y. Gao, R.-H. Li, J. Huang, J. Guo, L.A. Gutierrez, X. Wu, A fast parallel community discovery model on complex networks through approximate optimization, IEEE Trans. Knowl. Data Eng. 30 (9) (2018) 1638–1651.
[14] G.K. Zipf, The p1p2/d hypothesis: On the intercity movement of persons, Amer. Soc. Rev. 11 (6) (1946) 677–686.
[15] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, A tale of many cities: universal patterns in human urban mobility, PLoS One 7 (5) (2012) e37027.
[16] F. Simini, M.C. González, A. Maritan, A.-L. Barabási, A universal model for mobility and migration patterns, Nature 484 (7392) (2012) 96–100.
[17] F. Simini, A. Maritan, Z. Néda, Human mobility in a continuum approach, PLoS One 8 (3) (2013) e60069.
[18] Y. Yang, C. Herrera, N. Eagle, M.C. González, Limits of predictability in commuting flows in the absence of data for calibration, Sci. Rep. 4 (2014) 5662.
[19] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (7075) (2006) 462–465.
[20] M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782.
[21] H. Cai, X. Zhan, J. Zhu, X. Jia, A.S. Chiu, M. Xu, Understanding taxi travel patterns, Physica A 457 (2016) 590–597.
[22] M.S. Wang, L. Huang, X.Y. Yan, Exploring the mobility patterns of public transport passengers, J. Univ. Electron. Sci. Technol. China 41 (1) (2012) 2–7.
[23] C. Roth, S.M. Kang, M. Batty, M. Barthélemy, Structure of urban movements: polycentric activity and entangled hierarchical flows, PLoS One 6 (1) (2011) e15923.
[24] I. Rhee, M. Shin, S. Hong, K. Lee, S.J. Kim, S. Chong, On the levy-walk nature of human mobility, IEEE/ACM Trans. Netw. 19 (3) (2011) 630–643.
[25] W.H.K. Tsui, H.O. Balli, A. Gilbey, H. Gow, Forecasting of hong kong airport's passenger throughput, Tour. Manag. 42 (42) (2014) 62–76.
[26] Z.D. Zhang, J.H. Xu, An analysis of major factors on airport passenger volumes, Urban Transp. China 5 (6) (2007) 54–57.
[27] T. Jia, K. Qin, J. Shan, An exploratory analysis on the evolution of the us airport network, Physica A 413 (413) (2014) 266–279.
[28] F.H. Huang, J. Peng, M.Y. You, Analyses of characetristics of air passenger group mobility behaviors, Acta Phys. Sin. 65 (22) (2016) 228901.
[29] X.Y. Yan, C. Zhao, Y. Fan, Z. Di, W.X. Wang, Universal predictability of mobility patterns in cities, J. R. Soc. Interface 11 (100) (2013) 20140834.
[30] X. Liang, J. Zhao, K. Xu, A universal law in human mobility, arXiv preprint arXiv:1401.3918.
[31] Y.M. Zhao, A. Zeng, X.Y. Yan, W.X. Wang, Y.C. Lai, Universal underpinning of human mobility in the real world and cyberspace, Physics 18 (5) (2015) 053025.
[32] X.-Y. Yan, W.-X. Wang, Z.-Y. Gao, Y.-C. Lai, Universal model of individual and population mobility on diverse spatial scales, Nature Commun. 8 (1) (2017) 1639–1648.
[33] M. Shang, G. Chen, S. Dai, Interest-driven model for human dynamics, Chin. Phys. Lett. 27 (4) (2010) 048701.
[34] J. Alstott, E. Bullmore, D. Plenz, Powerlaw: a python package for analysis of heavy-tailed distributions, PLoS One 9 (1) (2013) e85777.
[35] J. He, X. Li, L. Liao, D. Song, W.K. Cheung, Inferring a personalized next point-of-interest recommendation model with latent behavior patterns, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 137–143.
[36] L. Minjin, H. Petter, Relating land use and human intra-city mobility, PLoS One 10 (10) (2015) e0140152.
[37] G.W. Harrison, Expected utility theory and the experiments, Empir. Econ. 19 (2) (1994) 223–253.
[38] M. Lenormand, S. Huet, F. Gargiulo, G. Deffuant, A universal model of commuting networks, PLoS One 7 (7) (2012) 457–464.
[39] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons, Biol. Skr. 5 (1948) 1–34.
[40] M. Ye, P. Yin, W.-C. Lee, D.-L. Lee, Exploiting geographical influence for collaborative point-of-interest recommendation, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2011, pp. 325–334.
[41] J.-D. Zhang, C.-Y. Chow, Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 443–452.
[42] P.A. Grabowicz, J.J. Ramasco, B. Gonçalves, V.M. Eguíluz, Entangling mobility and interactions in social media, PLoS One 9 (3) (2014) e92196.
[43] T. Zhou, X.P. Han, X.Y. Yan, Z.M. Yang, Z.D. Zhao, B.H. Wang, Statistical mechanics on temporal and spatial activities of human, J. Univ. Electron. Sci. Technol. China 6 (4) (2013) 481–540.