A Data-driven Approach to Design Feeder Bus Network based on Aggregated Cellphone Data and Open GIS Tool

Yaojun Wang[1]; Yue Liu (Corresponding Author)[2];

Abstract

This paper presents a mathematical model to design feeder bus network to access an existing urban rail system. A method was introduced to obtain real distribution of passenger demand mined from cellular data and travel impedance matrix calculated from an Open GIS tool. The proposed model features an integrated operational framework, which is able to simultaneously select bus stops, and dispatch and route buses from those targeted stops to urban rail stops. This paper further presents an improved GA–based heuristic approach to yield acceptable solutions to the model in a reasonable amount of time. The model is applied to a real-world case which aims to design a feeder bus system for Jiandingpo Station in Chongqin, China. More than 3.51×10^{8} cellular records were filtered and aggregated to obtain the associated demand patterns, and more than 2500 pairs of walking distances, travel time and vehicle distance between demand points and candidate bus stops, among candidate bus stops, were calculated with Open GIS tools to reflect real traffic status and network topology within study areas. Sensitivity analyses were also performed to investigate the impact of the number of designed bus routes on the model performance. The clarity of model inputs and its seamless integration with the commonly used Open GIS offer its best potential to be used as an effective tool for transit authorities to design and refine feeder bus network.

Keywords：Feeder bus network design, Cellphone data, Open GIS

1. Introduction

Over the past decades, the development of rail transit keeps blooming and benefiting the population in both urban and suburban area. Especially, it is widely recognized as a competitive mode to shift transport demand from individual car traffic to public transport and further enhance urban sustainability. However, the difficulty in conveniently transporting people from their home address to urban rail transit stations still exists. Such predicament in transportation is described as “Last Mile” problem. Since solving this issue has already been clearly recognized as a key step to greatly improve the usage of urban rail transit, many scholars have been making efforts and contributions to this area. Among those efforts, the feeder bus, an integration of urban rail transit and bus network, is now widely regarded as an effective tool to improve the service efficiency and financial status (Stanger and Vuchic, 1979; Dunn, 1980; May, 1991; Deng et al., 2013).

[1] Department of Civil and Environmental Engineering, University of Wisconsin at Milwaukee, P.O. Box 784, Milwaukee, WI 53201-0784, Tel: 414-334-9871, Email: yaojun@uwm.edu

[2] Department of Civil and Environmental Engineering, University of Wisconsin at Milwaukee, P.O. Box 784, Milwaukee, WI 53201-0784, Tel: 414-229-3857, Email: liu28@uwm.edu.

2. Literature Review

Existing research focuses on Feeder Bus Network Design Problem (FBNDP), mainly falling into two categories (Deng et al., 2013), analytic approach and network programming. The analytic approach derives the optimal route spacing, operating headway and the optimal stop spacing. A couple of pioneering studies were proposed to optimize the location, headway of feeder buses as well as determining the number and the length of routes (Byrne and Vuchic, 1972; Hurdle, 1973; Byrne, 1976; Chien and Yang, 2000). Wirasinghe (1980) presented an approximate analytical model and corresponding solution algorithm to design a feeder bus system access to a rail station served a peak-period demand. The model was further applied to the Calgary (South Corridor) LRT system. In addition to optimizing the route spacing, operating headway and the stop spacing, Kuah and Perl (1988) also discussed the influencing factors of bus-stop spacing in three different cases. Chowdhury (2002) took stochastic feeder vehicle arrivals at transfer stations into account, and then the slack times of coordinated routes were optimized by balancing the savings from transfer delays and additional cost from slack delays and operating costs. Although the analytic models have been widely used to deal with FBNDP, their basic assumptions, the shape of the street geometry and the spatial distribution of demand, are regarded as the significant limitations (Deng et al., 2013).

In recent years, an ever-increasing interests of using the network programming approach to handle FBNDP have been raising. Typically, this approach decomposes the traffic network into a set of nodes and a set of links. Some of nodes are selected as bus stops while some of them represent urban rail transit stations. The links between stops, and between stop and station are treated as bus route segments. The demand is centered on the nodes and transported from node to node. Many mathematical programming models, especially, integer and mixed integer programs, are developed to handle FBNDP under either M-to-1 or M-to-M demand pattern (Kuah and Perl, 1989; Martins and Pato, 1998; Gholami, 2011). Based on those developed methodologies, some more advanced studies are followed. Ciaffi (2012) proposed a two-phase mode to cope with FBNDP. At the first phase, a heuristic algorithm is developed to generate two different and complementary sets of feasible routes and then, with the output from the first phase, a GA-based algorithm is utilized for finding a sub-optimal set of routes with the associated frequencies at second phase. Considering a multi-level cost structure, including, passengers' cost and operators' cost, Deng et al. (2013) proposed a model to solve the issue of M-to-M feeder bus network. Pan (2014) employed a bi-level model to maximize the number of served passengers by the feeder transit system in the upper level and to minimize the operational cost for transit operators in the lower lever level.

Although most of the aforementioned studies have successfully handled a variety of FBNDPs, the following critical issues deserve further investigation during the process of design feeder bus network:

1. Most studies have neglected the integrated operation of pedestrian guidance (from home addresses to candidate bus stops) and transit routing (from selected bus stops to urban rail transit stop). Since the stops are the first point to connect bus services with passengers, the ignorance of passengers' interests easily results in decreasing the attractiveness of transit riders (Lin et al., 1995; Fu et al., 2003; Sun and Hickman, 2005; Chien et al., 2010).

2. Most of existing studies have set the shape of the street geometry and spatial distribution of demand as the basic assumptions for model development. The first assumption is unable to reflect the characteristic of real traffic network, such as, one-way street or left-turn only intersection. Considering realistic network has been widely regarded as a vital role in bus network design problem as an accurate traffic network contributes to the reliability of the result (Chien, 2003; DiJoseph and Chien, 2013). And the second one directly leads to an ambiguity in determining the

location of bus stops. Both of them could make the case study fail in practicability and authenticity.

3. The real traffic status, especially the travel time which changes dramatically under different network conditions, likely, peak period, non-peak period and weather condition, has not been included when proposing methodology for FBNDP.

3. Research Motivation

Realizing above reviewed limitations of existing studies on FBNDP, the paper will focus on the following critical research tasks:

1. Propose an integrated optimization model that is capable of seamlessly and simultaneously coordinating the passenger boarding guidance and transit routing process when development of feeder bus network;

2. Develop a cellphone data processing methodology that can extract the spatial distribution of urban rail transit user demand to cope with demand uncertainty issue in existing studies;

3. Introduce an Open-GIS tool (Baidu Map) to retrieve the features and travel time information of real traffic network which is able to reflect the network topology and real traffic status;

4. Develop a heuristic solution algorithm to efficiently yield the acceptable solution to the proposed model;

5. Illustrate the proposed methodology through a real-world case study to best understand and apply the proposed methodology during the design process of feeder bus system.

4. Methodology

4.1 Research Framework

In this paper, a feeder bus system is proposed that provides services to conveniently transport passengers to urban rail transit station. Such a system benefits from using cellphone data to extract real distribution of passenger demand which is expected to efficiently solve the issue of demand uncertainty in traditional feeder bus design problem. Moreover, an open GIS tool is further developed to offer traffic status and topological features within study area when designing feeder bus network. With aggregated demand patterns, traffic and network information, a mixed-integer programming is formulated to select the most appropriate locations of bus stops, and guide passengers from demand points to their associated stops. Then the model dispatches and routes buses from selected stops to urban rail transit station. Those key components can be better illustrated by a research framework graph, as shown in Figure 1.

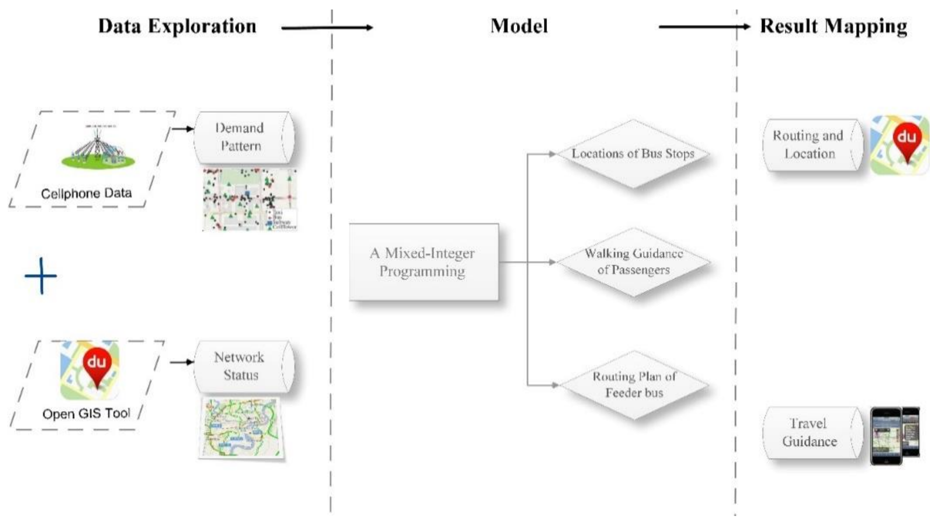

Fig. 1 Research Framework

In Figure 2, five candidates (red dots) are finally selected as feeder bus stops due to a high accessibility to surrounding demands points (black dots). Three example bus routes are highlighted and shown in solid red line (node 1-node 2- urban rail transit station, node 3-node 4- urban rail transit station and node 5-node 6- urban rail transit station). The aim of the proposed mixed integer program is to find a sub-graph that simultaneously minimize weighted passengers walking distance and operational cost of feeder bus system. In this illustrated network, passengers are assigned to stops with respect to minimizing their walking distances that is also the principle of determining the locations of bus stops in our study. Once the demand is assigned at the stops, the feeder route for each bus will be constructed to transfer passengers waiting at the stops to their connected urban rail transit station by finding out the shortest path. Different from phase-based or stage-based approaches in existing studies, we formulate it as a combined stop location and vehicle routing problem.

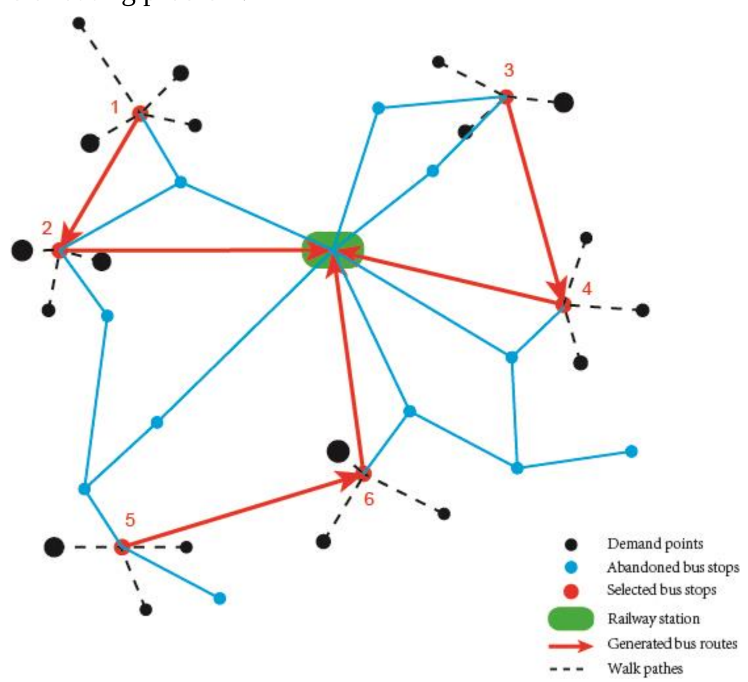

Fig. 2 Graphical representation of the integrated FBDNP problem

4.2 Data Processing
4.2.1 Cellphone Dataset

In this section, we describe a cellphone data processing for extracting demand distribution of urban rail transit user at morning peak. Over the last decade, the massive cellphone data has already been widely proved to be a promising data source for estimation of aggregate level origin-destination matrix (Joanna and Ivan, 2002; Caceres et al., 2007; Francesco, et al., 2011). In principle, when a user initiates a communicating event (e.g. a voice-call or text), a cell tower which contains location information needs to assigned to serve this event. Thus, a corresponding positioning record (containing information on users' locations) is subsequently generated as the event occurs. Such record is further utilized for estimating personal travel trajectory at aggregated level. Mobile phone data used in this paper was collected by the base transceiver stations (BTS) that are used as treated as the fixed traffic detector to monitor the movement of mobile users. It should be noted that the communication record (associating with commutating event) typically is not directly containing geo-location information but only providing a unique ID of the communicating tower (a combination of LACID and CellID). Therefore, a matching job between communication records and BTS geo-location table is required to obtain positioning information for each record. After the matching, the data format of communication records used in this study is shown in table 1.

Table 1 Format of communication record

Fields	Description
MSID	Identity of a mobile user
MSTIME	Timestamp of the acquired communicating event
LACID	ID of location area code of connected BTS
CellID	ID of cell identifier of the connected BTS
EventID	ID of the acquired communication event
LON	Longitude of the geographic coordinates of the connected BTS
LAT	Latitude in the geographic coordinates of the connected BTS

The procedure and the methods of processing cellphone dataset is outlined in Figure 3.

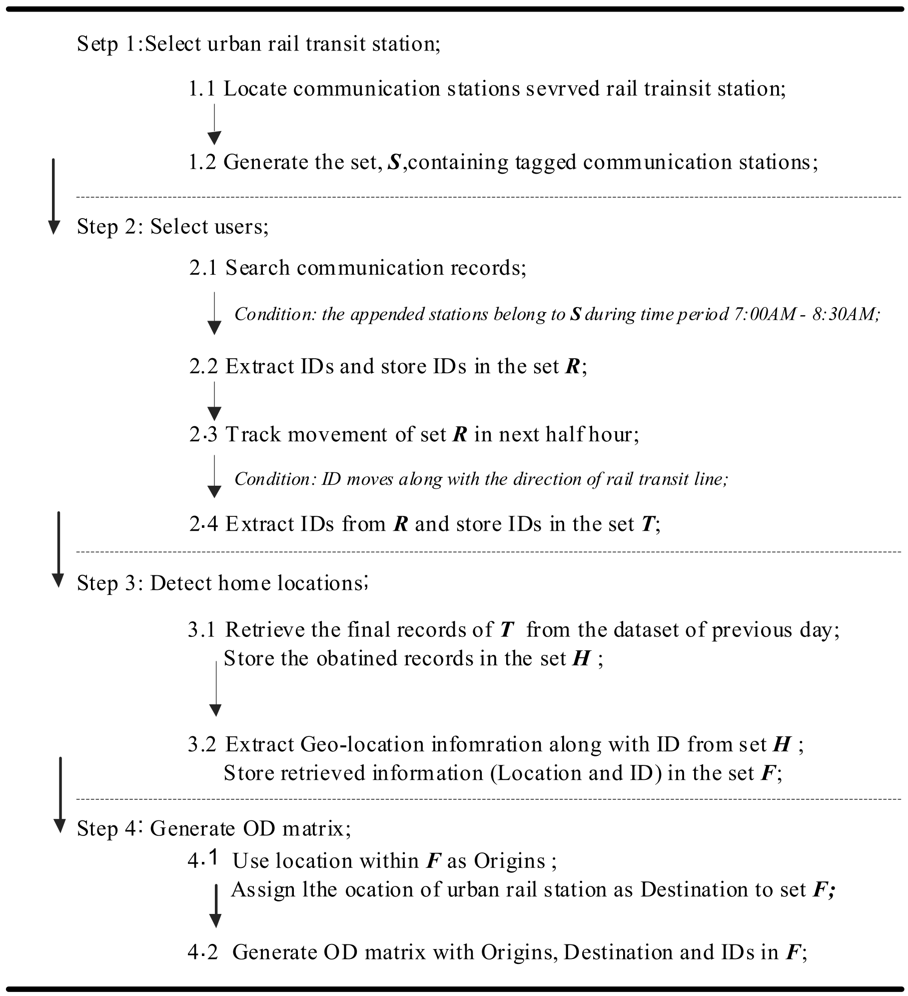

Fig. 3 The OD generation procedures

4.2.2 Open GIS Dataset

Nowadays, almost all of popular map-engine providers, such as Google Map, Openstreet and Baidu Map etc., offer a full access to their GIS database by calling defined Application Program Interface (API). Such application enables us to achieve various kinds of information, namely, geo-location, transportation and etc. In our case, we concentrate in retrieving map-based travel distance and time between involved nodes. For example, in Google Map APIs, one can easily requests the distance matrix data between Washington, DC and New York City, NY by calling https://maps.googleapis.com/maps/api/distancematrix/json?units=imperial&origins=Washington,DC&destinations=New+York+City,NY&key=YOUR_API_KEY" in JASON, JavaScript or Python format.

In this study, a scripting language, Java Script, is used to develop the interactive tool connecting with Open GIS engine which aims to collect map-based travel distance and time data based on input OD pairs.

4.3 Model Formulation

4.3.1 Notation

To facilitate the model presentation, all definitions and notations used hereafter are summarized in the table 2.

Table 2 Parameters and variables in the mathematical model

Indices	
i	Demand point index
j,m,p	Vehicular node (bus stop candidates and urban rail transit station) index
k	Bus route index
Sets	
$\boldsymbol{I}$	Set of demand points
$\boldsymbol{M}$	Set of bus stop candidates
$\boldsymbol{MS}$	Set of urban rail transit stations, without loss of generality we are assuming a single station in this model
$\boldsymbol{K}$	Set of bus routes
Parameters	
$Demand_i$	Number of passengers at demand point i; $i \in I$
P	Maximum number of designed bus stops;
Q_k	Capacity of bus route k; $k \in K$
T_{max}	Maximum travel time;
L_{min}	Minimum route length;
d_{ij}	Map-based walking distance from demand point i to bus stop candidate j; $i \in I, j \in M$
t_{jm}	Map-based vehicular distance from vehicular node j to vehicular node m; $j, m \in M \cup MS$
C_m	Operational cost per km (unit: dollar);
C_h	Operational cost for drivers per operating hour (unit: dollar);
C_p	The value of passenger's walking time per hour (unit: dollar);
Decision Variables	
c_{jm}^k	Number of passengers at stop j assigned to route k traveling from j to m (unit: person);
U_{ik}	An auxiliary (real) variable for sub-tour elimination constraint in route of bus k;
$z_{jm}^k = \begin{cases} 1 \\ 0 \end{cases}$	If stop j precedes point m on the route k; Otherwise
$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}$	If demand point i is assigned to stop candidate j; Otherwise;
$y_j = \begin{cases} 1 \\ 0 \end{cases}$	If candidate node j is selected as a stop; Otherwise;

4.3.2 Formulation

The proposed problem can be formulated as the following mixed integer program (MIP):

Minimize:

$$(C_m \sum_{j\in M} \sum_{m\in M\cup MS} \sum_{k\in K} Z_{jm}^k d_{jm} + C_h \sum_{j\in M} \sum_{m\in M\cup MS} \sum_{k\in K} Z_{jm}^k t_{jm}) + C_p \sum_{i\in I} \sum_{j\in M} demand_i \, d_{ij} x_{ij} \quad (1)$$

Subject to:

$$\sum_{j\in M} y_j \le P; \quad (2)$$

$$x_{ij} \le y_j \qquad \forall i \in I, \forall j \in M; \quad (3)$$

$$\sum_{j\in M} x_{ij} = 1 \qquad \forall i \in I; \quad (4)$$

$$2 * z_{jm}^k \le y_j + y_m \qquad \forall k \in K, \forall j \in M, \forall m \in M \cup MS; \quad (5)$$

$$\sum_{j\in M} \sum_{k\in K} z_{jm}^k \le 1; \quad \forall m \in M; \quad (6)$$

$$\sum_{p\in M} \sum_{k\in K} z_{pj}^k \le 1; \quad \forall j \in M; \quad (7)$$

$$\sum_{m\in M\cup MS} z_{jm}^k - \sum_{p\in M} z_{pj}^k \ge 0; \qquad \forall j \in M, \forall k \in K; \quad (8)$$

$$U_{ik} - U_{jk} + \left(|H| * z_{jm}^k\right) \ge |H| - 1, \forall j, m \in M \cup MS, \forall k \in K; \quad (9)$$

$$\sum_{j\in M} \sum_{m\in M\cup MS} c_{jm}^k \le Q_k; \qquad \forall k \in K; \quad (10)$$

$$\sum_{m\in M\cup MS} \sum_{k\in K} c_{jm}^k = \sum_{i\in I} Demand_i * x_{ij}; \qquad \forall j \in M; \quad (11)$$

$$c_{jm}^k - z_{jm}^k \ge 0; \qquad \forall j \in M, \forall m \in M \cup MS, \forall k \in K; \quad (12)$$

$$c_{jm}^k \le Q_k * z_{jm}^k; \qquad \forall j \in M, \forall m \in M \cup MS, \forall k \in K; \quad (13)$$

$$\sum_{j\in M} \sum_{m\in M\cup MS} z_{jm}^k * t_{jm} \le T_{max}; \ \forall k \in K; \quad (14)$$

$$\sum_{j\in M} \sum_{m\in M\cup MS} z_{jm}^k * d_{jm} \ge L_{min}; \ \forall k \in K; \quad (15)$$

$$\sum_{j\in M} \sum_{m\in MS} z_{jm}^k = 1; \qquad \forall k \in K; \quad (16)$$

$$\sum_{j\in M} \sum_{m\in MS} z_{mj}^k \le 0; \qquad \forall k \in K; \quad (17)$$

In this formulation, the objective function is given by Eq. (1), which includes two terms: the first term is dealing with routing and the second one is related to location selection of bus stops and assigning passengers to those targeted stops. The first term minimizes the total operational cost for designed feeder bus system and the second term minimizes the total equal value of the walking distances from demand points to selected stops.

Constraint (2) indicates that the number of selected bus stops should be no more than the allowed maximum number. Constraint (3) and (4) guarantee the demand points can only be assigned to those selected stop candidates and each demand point must be matched and served by only one bus stop. Constraint (5) specify the bus route links may exist between two candidate nodes only if both candidates are selected as stops. Constraint (6) guarantees each selected bus stop can only serve one bus route which aims to avoid undesirable competition among bus routes and further increase the whole system efficiency. Constraint (7) and (8) set each bus stop (except urban rail station) being served to have exactly the same incoming and outgoing arcs. Constraint (9) is used for sub-tour elimination in the vehicle routing problem and is a constraint with polynomial cardinality (Miller, 1995). Constraint (10) guarantees the number of passengers in each route boarding from selected bus stops and transported to the urban rail transit station must be less than the vehicle capacity during each route. Constraint (11) and (12) ensure that all the passengers are picked up. Constraint (13) guarantees that passengers are assigned to the route only if this route server that link or selected stop candidates. Constraint (14) and (15) are used to limit the minimum length and maximum travel time for each route. Constraint (16) and (17) ensure that each route is eventually ended at the urban rail transit station.

4.4 A GA-based Heuristic Algorithm

The proposed optimization model is an extension to the vehicle routing problem that has been proved to be non-deterministic polynomial-time hard (NP-hard). In case of small-scale networks, some powerful solvers such as IBM ILOG CPLEX may be able to better find an optimal solution for proposed problems. However, those tools are intractable for those large-scale network problems. Thus, a GA-based heuristic approach is further developed to efficiently yield acceptable solutions to the model in a reasonable amount of running time.

4.4.1 Coding of GA chromosomes

An efficient coding of GA chromosomes which is able to capture the characteristics of the solution structure plays a key role in the process of GA searching. In our study, the main body of proposed optimization model is composed of selecting stops locations, matching demand points with candidate bus stops, and designing routing plans that are corresponding to y_j, x_{ij} and z_{jm}^k respectively. Therefore, if we use vectors $U = (u_1, u_2, \dots, u_M, u_{M+1}, \dots, u_{2M}, u_{2M+1}, \dots, u_{2M+I})$ to represent solutions to this model, each vector consists of $(2M + I)$ binary strings, it could be further decomposed into three parts, and explained as following:

1. The first part of GA chromosomes $(u_1, u_2, \dots, u_M)$ (the vector of binary variables) is used to represent the decision of location selection of bus stops. If $u_j = 1$, then the corresponding candidate nodes *j* is targeted as a feeder bus stop;
2. The second part of GA chromosomes $(u_{M+1}, u_{M+2}, \dots, u_{2M})$ (the vector of integer variables) is used to assign the selected stops to different routes, thus each *u* ranges from 1 to k where $k \in K$ represents the number of designed bus routes. For example, $u_{M+2} = 3$ indicates that bus stop *(M+2)* is assigned to route 3. And then a Dijkstra algorithm is implemented to search the shortest bus route dispatching from urban rail transit station so as to order the sequence of targeted stops for each bus route;
3. The third part of GA chromosomes $(u_{2M+1}, u_{2M+2}, \dots, u_{2M+I})$ (the vector of integer variables) is used to match demand points to the closest stops for each bus route. Each *u* ranges from 1 to *k* where $k \in K$ represents the number of designed bus routes. The function of $(\sum_{i\in I}\sum_{j\in M} demand_i\, d_{ij}x_{ij})$ in model objective is utilized to obtain the distance matrix so as to determine value of $x_{ij} = \{i | min \sum_i demand_i d_{ij}\}$ which help each demand point find it's most appropriate boarding stop.

4.4.2 Fitness Evaluation

Note that the candidate solutions may violate constraints (10), (14) and (15). To deal with this problem, we include those constraints as penalty terms into the function of fitness evaluation. Thus, the modified fitness function in our study is given by:

$$F = f + M_1 \cdot \sum_{k\in K}\left(\max\left(\sum_{j\in M}\sum_{m\in M\cup MS} c_{jm}^k - Q_k, 0\right)\right)^2 +$$
$$M_2 \cdot \sum_{k\in K}\left(\max\left(\sum_{j\in M}\sum_{m\in M\cup MS} z_{jm}^k * t_{jm} \leq T_{max}, 0\right)\right)^2 +$$
$$M_3 \cdot \sum_{k\in K}\left(\max\left(\sum_{j\in M}\sum_{m\in M\cup MS} z_{jm}^k * d_{jm} \geq L_{min}, 0\right)\right)^2$$

where f is the objective function (Equation 1) of the proposed model; F is the function used in fitness evaluation. And M_1, M_2, M_3 are large positive penalty coefficients.

4.4.3 A Heuristic Algorithm of Generating Initial Population

As it has been widely recognized, the quality of the solution found, or the computational resources required by applying GA-based algorithm, highly depends on the selection of initial population to the proposed problem. Thus, to better solve the presented model and improve computational efficiency, a heuristic algorithm for generating feasible initial population is further developed to embed into GA process. The procedures are explained as following:

Step 1. Input parameters defined within the proposed model, namely, M (a set of candidate bus stops), MS (urban rail transit station) and K (a set of feeder bus routes) etc.;

Step 2. For each route $k \in K$, starting with node of rail station: 1) initial a feasible set of candidate bus stops $M' \in M$ in which distance between selected node and rail station is less than L_{min}; 2) randomly select a candidate bus stop from M'; 3) repeat searching next node from the rest nodes in M' until violating the constraints (2), (14) and (15);

Step 3. Set $y_j = y_m = 1$ in case of $z_{jm}^k = 1$ for each route $k \in K$;

Step 4. Calculate the distance between each demand point $i \in I$ and all confirmed bus stops, and then determine the values of x_{ij}which is for assigning the most appropriate bus stop to each demand point with considering the capacity constraints and minimum walking distance;

Step 5. Use the obtained decision of $y_{j,} x_{ij}, z_{jm}^k$ from Step 1 to 4 to generate initial population U;

4.5 Genetic Operators

4.5.1 Selection

Selection operators give preference to better solutions (chromosomes), allowing them to pass on their 'genes' to the next generation of the algorithm. In our study, uses both of random competition and elitist selection strategies to ensure that individuals with highest fitness in the previous population are retained in the next population.

4.5.2 Crossover and Mutation

Crossover operator simulates recombination for exchange part of genes in two individuals to produce new individuals in evolutionary processes. In our study, we use one-point method which randomly select an integer P between 1 and $(2M + I)$, and exchange the front and the rear portions of two parent U_1 *and* U_2 to generate new offspring chromosome $U_1{}'$ *and* $U_2{}'$. Mutation operator in this study is also implemented with one-point method where we define mutation fraction to 0.15. If gene $u_j \in U$ has been selected as a mutation point, and then u_j is set to 1 or 0 in case of $j \in [1, M]$ while u_j randomly takes value range from 1 to $|K|$.

4.5.3 Stopping Criteria

The GA stops to evolve until the following criteria are met:

(1). $\left|\frac{\hat{\mathcal{F}}_{min}^{n} - \hat{\mathcal{F}}_{min}^{n-1}}{\hat{\mathcal{F}}_{min}^{n}}\right| < \epsilon$, i.e., the difference between the minimum evaluation values between two adjacent generations is less than a threshold ϵ; or

(2). A pre-set maximal number of generations are reached.

5. Case Study

To illustrate the applicability of the proposed Data-driven framework and models in designing feeder bus network access to the urban rail transit, this study has selected Jiandingpo Station at Metro Line 1 in Chongqing (the biggest municipality under direct administration by the Chinese central government), for a case study.

As the first stop of Line 1, Jiandingpo station is located at west part of Chongqing. Its geographical location is given by Figure 4

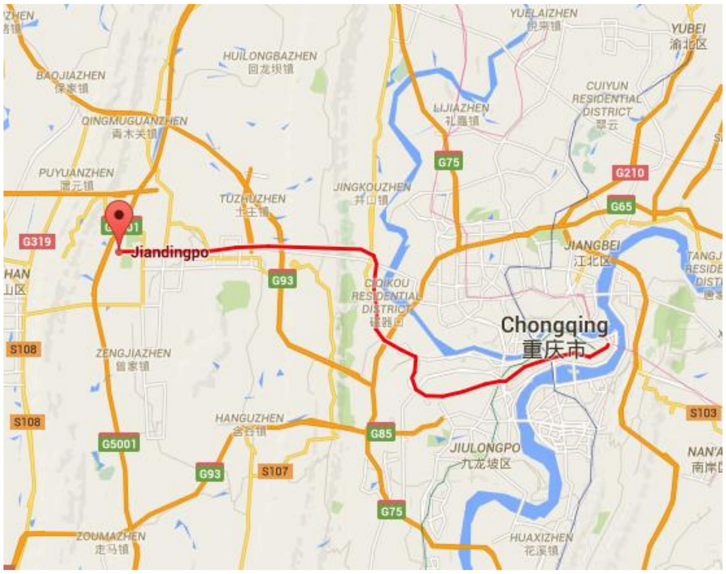

Fig. 4 GeoLocation of Jiandingpo Station (Source: Google Map)

5.1 Cellphone Dataset and Open GIS Dataset

5.1.1 Cellphone Dataset

The cellphone dataset used in this case study to extract the spatial distribution of demand in relation to Jiandingpo station is collected from both China Mobile and China Unicom (the two biggest communication operators in China) during March 4th 2015 and March 5th 2015. Table 3 descripts some basic information about dataset used in this study.

Table 3 Cellphone Dataset Description

Day	Size	No. Records	No. Detected Devices
March 4th 2015	22.5GB	1.78×10^8	7.61×10^6
March 5th 2015	21.9GB	1.73×10^8	7.58×10^6

By applying the data processing methodology explained in section of "Methodology", 25 demand points containing 513 passengers who catch Line 1 at Jiandingpo Station in morning peak hour are located at the map. Moreover, after a detection of demand distribution, we pick up 42 candidate bus stops surrounding those 25 demand points based on local traffic network. Figure 5 is used to map both inferred demand points and chose candidate stops where larger size of blue dot represents a larger demand size and red dots represent candidate bus stops. Thus, the main objective of this case study is to design a convenient feeder system for transporting those detected demand to Jiandingpo Station at morning peak hour.

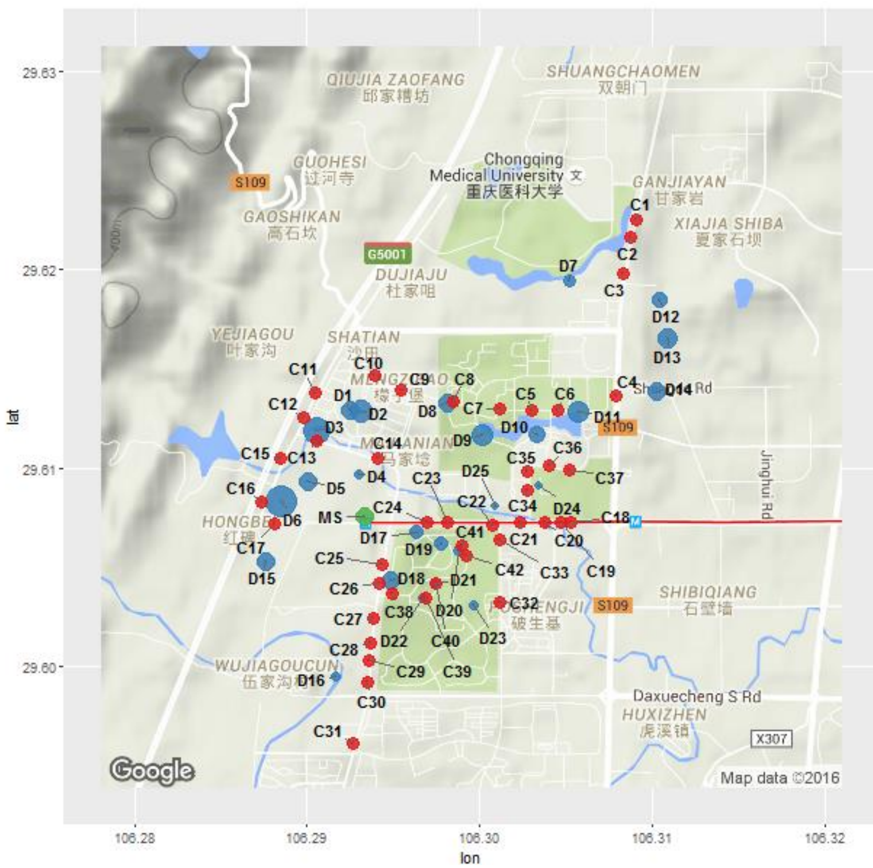

Fig. 5 Spatial distribution of demand using Jiandingpo Station and candidate bus stops (Map resource: Google)

The number of passengers corresponding to demand points is also recorded in table 4.

Table 4 No. Passengers corresponding to demand points

Demand Point	No. Passengers	Demand Point	No. Passengers
D1	23	D16	6
D2	36	D17	12
D3	50	D18	25
D4	7	D19	12
D5	22	D20	8
D6	70	D21	6
D7	11	D22	7
D8	25	D23	6
D9	32	D24	5
D10	21	D25	5
D11	30	D14	24
D12	17	D15	25
D13	28		

5.1.2 Open GIS Dataset

The Baidu Map APIs (similar to Google Map APIs) is adopted here to offer both traffic status and network information, specifically, travel time during and the shortest path choice during morning peak hour. Specifically, about 1050 pairs (25 demand points x 42 candidate bus stops) of walking distance between demand points and candidate bus stops are extracted while another 1849 ((1 station + 42 candidates) x ((1 station + 42 candidates)) pairs of shortest travel distance and time during morning peak hour among candidate bus stops and between candidate stops and urban trail transit station are also obtained.

5.2 Results Analysis

Key parameters used in the case study are given as follows:

- No. Bus routes: 3;
- Route capacity: 200 persons;
- Maximum allowed No. Stops: 16;
- Maximum allowed travel time for each route: 20 min;
- Minimum route length: 2 km;
- Operational cost for feeder buses: $3 per km;
- Operational cost for drivers: $5 per hour.
- The value of passenger's walking time: $1 per hour

Using the demand collected from cellphone data exploration, distance and time matrix generated from Open GIS tool, the proposed model was firstly solved in *CPLEX 12.6* to optimality. Table 5 summarizes the assignment results of passengers from demand points to selected candidate bus stops which is also the decision of location selection of bus stops.

Table 5 Assignment of passengers from demand points to selected bus stops

Demand Point	**Candidate Bus Stop**	**Served Demand (Unit: person)**	**Bus Route**	**Walking Distance (Unit:m)**
D7	C3	11	R1	1417
D12	C3	17		209
D13	C4	28		255
D14	C4	24		350
D11	C6	30		434
D23	C32	6		206
D24	C34	5		93
D25	C34	5		261
D10	C36	21		283
D17	C41	12		293
D19	C41	12		211
D20	C41	8		16
D9	C8	32	R2	267
D1	C10	23		360
D2	C10	36		344
D3	C13	50		8
D8	C8	25		62
D4	C14	7		462
D5	C16	22	R3	187
D6	C16	70		258
D15	C16	25		272
D16	C30	6		242
D18	C38	25		95
D22	C39	7		13
D21	C40	6		10

Table 6 details the routing plans for each bus route. Due to the fact that the inputs are generated from Open-GIS tool, the result of vehicle travel time/distance and walking distance are capable of representing the real traffic status and network topology within study area. A map-based graphical illustration of bus routing plans as well as passenger guidance is shown in Figure 6. The red solid line represents Route 1, blue solid line represents Route 2, and black dash line represents Route3. In addition to bus route plan, those green solid lines are used indicate the walking paths of passengers.

Table 6 Routing Plans and Number of Passengers Served

Route Index	**Routes**	**Length(km)**	**Travel Time(min)**	**Served Demand**	**Weighted Average Walking Distance(m)**
1	C3-C4-C6-C36-C34-C32-C41-MS	5.07	15.5	179	351
2	C10-C8-C14-C13-C16-MS	3.90	11.5	290	219
3	C30-C39-C40-C38-MS	2.05	6.0	44	90

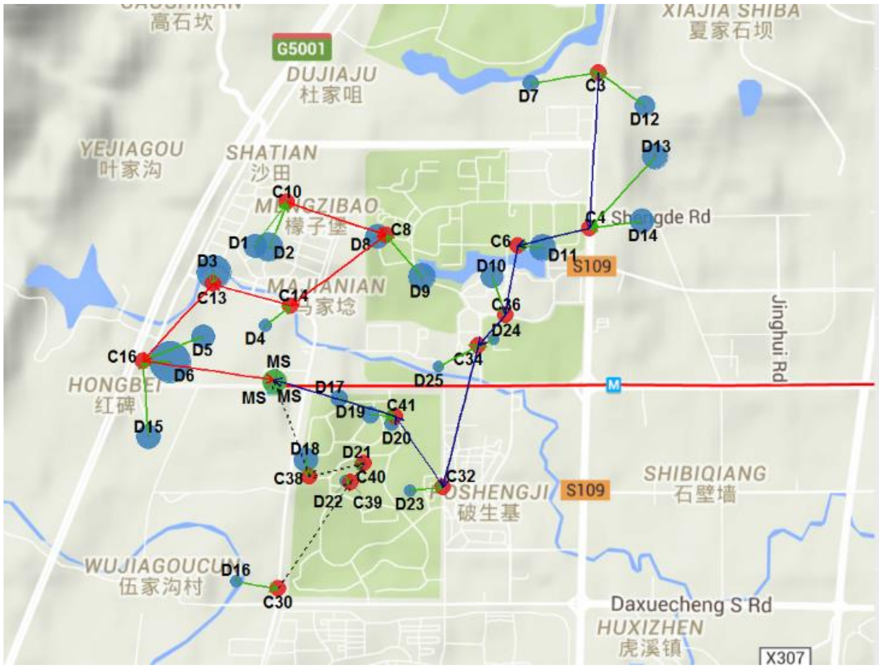

Fig. 6 Case Study Result (Map source: Google)

CPLEX 12.6 has successfully found a global optimization to this case. However, it takes about 390 seconds for model to compute, which may raise an issue of computation efficiency and hinder its application in large-scale or more complicate cases. Thus, we also implement the GA-based heuristic algorithm developed in this paper to solve the scenario of 3 routes, 4 routes and 5 routes respectively. Figure 7 shows the convergence process of GA algorithm for three scenarios. As more routes are designed, more iterations are required to convergence. A result comparison of CPLEX and GA algorithm is unfolded in terms of computation efficiency and solution difference, as recorded in Table 7. One can observe that the computation time of CPLEX solved for global optimization is up to more than one hour in case of designing 5 routes while the developed heuristic algorithm takes about 82 seconds to find an acceptable near optimal solutions. The error of average bus route length and travel time is controlled under 20% only except for scenario of 4 routes problem. The difference of weighted average walking distance is varied from 15% to 30% which increases along with the number of routes.

Table 7 Comparison of CPLEX Solution and Heuristic Solution

	CPLEX Results				Heuristic Results				Difference		
Scenario	Solved Time (s)	Average Route Length (km)	Average Travel Time (min)	Average Weighted Walking Distance (m)	Solved Time (s)	Average Route Length (km)	Average Travel Time (min)	Average Weighted Walking Distance	Average Route Length	Average Travel Time	Weighted Walking Distance
3 Routes	390	3.67	11.00	253.74	48	4.17	12.44	292.94	14%	13%	15%
4 Routes	573	2.61	7.85	253.74	66	3.15	9.63	307.13	21%	23%	21%
5 Routes	3813	2.29	6.72	253.74	82	2.64	7.89	329.59	15%	17%	30%

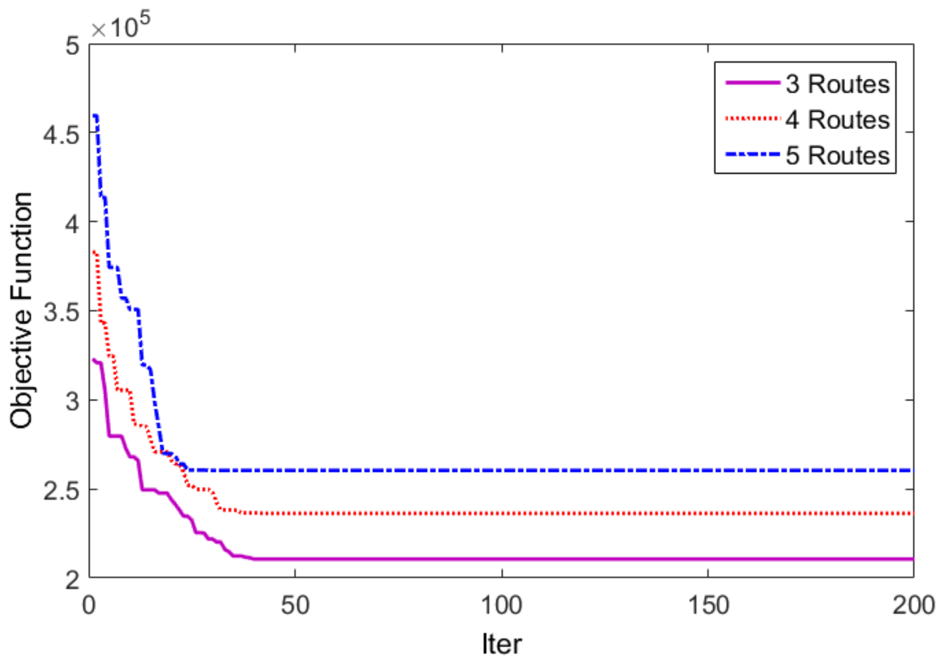

Fig. 7 Convergence Process of GA Algorithm of Three Scenarios

5. Conclusion

This paper presents a data-driven approach for designing feeder bus network connecting to the urban rail transit station. Different from existing studies, the proposed methodology features in: 1) developing a mixed integer programming to offer an interactive process of pedestrian guidance (from home addresses to candidate bus stops) and transit routing (from selected bus stops to urban rail transit stop, such integration will significantly improve the performance of the feeder bus system; 2) introducing a big data processing technology for extracting aggregated-level spatial distribution of demand with using cellphone dataset to solve the issue of demand uncertainty; 3) retrieving map-based travel distance and time information to include the network characteristics and traffic status by using Open GIS tool; 4) develop an improved GA-based heuristic algorithm in which a heuristic algorithm of generating the initial population is further proposed and embedded. The feasibility and applicability of the proposed model is illustrated with a real-world example, Jiandingpo Station of Chongqing Metro Line1, solved to optimality. Results show that the proposed model can yield valid and detailed passenger walking guidance and transit routing plans for feeder bus system. In order to validate the performance of developed heuristic algorithm, a comparison of CPLEX solutions and heuristic algorithm solutions in case of designing 3 routes, 4 routes and 5 routes respectively, the results suggest that the proposed algorithm is able to yield effective solutions to the proposed problem in an acceptable computation time.

Note that the problem studied in this paper is static in the way that the inferred OD table and the number of bus routes are all stable. Determination of bus stops locations and assignment of demands and routing also use a static representation of the network. Therefore, this model is very useful at the initial stage of strategic feeder bus network planning. Specially, because of a good connection with map engine, this model also has a potential to embed into transit APPs' development which is able to guide passengers starting trips from home. Extending the model to an explicitly dynamic setting with time-varying demand generation rates and travel times is a worthwhile direction for further work and future research.

Reference

Byrne B.F., Vuchic V. (1972). Public transportation line positions and headways for minimum cost. Traffic flow and transportation, 347-360.

Byrne B.F. (1976). Cost minimizing positions, lengths and headways for parallel public transit lines having different speeds. Transport Research, 10(3), 209-214.

Caceres, N., Wideberg, J. P., & Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. Intelligent Transport Systems, IET, 1(1), 15-26.

Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Computing, 10(4), 0036-44.

Chen, C., Zhang, D., Zhou, Z. H., Li, N., Atmaca, T., & Li, S. (2013, March). B-Planner: Night bus route planning using large-scale taxi GPS traces. In Pervasive Computing and Communications (PerCom), 2013 IEEE International Conference on (pp. 225-233). IEEE.

Chien, S., Tsai, F. M., & Hou, E. (2003). Optimization of multiple-route feeder bus service: application of geographic information systems. Transportation Research Record: Journal of the Transportation Research Board, (1857), 56-64.

Chien, S., & Yang, Z. (2000). Optimal feeder bus routes on irregular street networks. Journal of Advanced Transportation, 34(2), 213-248.

Chien, S., Byun, J., Bladikas, A., 2010. Optimal stop spacing and headway of congested transit system considering realistic wait times. Transportation Planning and Technology 33 (6), 495–513.

Chowdhury, S. M., & I-Jy Chien, S. (2002). Intermodal transit system coordination. Transportation Planning and Technology, 25(4), 257-287.

Ciaffi, F., Cipriani, E., & Petrelli, M. (2012). Feeder bus network design problem: a new metaheuristic procedure and real size applications. Procedia-Social and Behavioral Sciences, 54, 798-807.

Deng, L. B., Gao, W., Zhou, W. L., & Lai, T. Z. (2013). Optimal design of feeder-bus network related to urban rail line based on transfer system. Procedia-Social and Behavioral Sciences, 96, 2383-2394

Dunn J.A. (1980). Coordination of urban transit services: the German model. Transportation, 9, 33-43.

Fu, L.P., Liu, Q., Calamai, P., 2003. Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record 1857, 48–54.

Hurdle V.F. (1973). Minimum cost locations for parallel public transit lines. Transport Science, 7, 340 350.

Kuah G.K., Perl J. (1988). Optimization of feeder bus routes and bus-stop spacing. Transport Engineering, 114(3), 341-54.

Kuah GK, Perl J. (1989). The feeder-bus network-design problem. Journal of the Operational Research Society, 40, 751-767.

Matins C.L., Pato M.V. (1998). Search strategies for the feeder bus network design problem. Europe Journal Operational Research, 106, 425-440.

May A.D. (1991). Integrated transport strategies: a new approach to urban transport policy formulation in the UK. Transport Reviews, 2(3), 233-247.

Miller, D. L. (1995). A matching based exact algorithm for capacitated vehicle routing problems. ORSA Journal on Computing, 7(1), 1-9.

Pan, S., Yu, J., Yang, X., Liu, Y., & Zou, N. (2014). Designing a Flexible Feeder Transit System Serving Irregularly Shaped and Gated Communities: Determining Service Area and Feeder Route Planning. Journal of Urban Planning and Development, 141(3), 04014028.

Stanger R.M. & Vuchic V.R. (1979). The design of bus-rail transit facilities. Transit Journal, 5, 61-72.

Sun, A., Hickman, M., 2005. The real-time stop-skipping problem. Journal of Intelligent Transportation Systems 9 (2), 91–109.

White, J., & Wells, I. (2002). Extracting origin destination information from mobile phone data. In Road Transport Information and Control, 2002. Eleventh International Conference on (Conf. Publ. No. 486) (pp. 30-34). IET.

Wirasinghe S.C., Hurdle V.F., Newell G.F. (1977). Optimal parameters for a coordinated rail and bus transit system. Transport Science, 11(4), 359-374.