Dynamic Bus Route Adjustment Based on Hot Bus Stop Pair Extraction

Jiaye Liu, Jiali Mao ✉, YunTao Du, Lishen Zhao, and Zhao Zhang

School of Data Science and Engineering, East China Normal University, China
51184501030,10153903105,5118510035@stu.ecnu.edu.cn jlmao, zhzhang@dase.ecnu.edu.cn

Abstract. The crowdedness of buses caused by limited public transportation capacity has already severely influenced the convenience and comfort of inhabitant trip. Existing measures that reducing dispatching interval and replenishing more buses can soothe this case while aggravate traffic jam. To address the issue of inconvenient public transit characterized by packed buses, we propose a data-driven route adjustment framework, called Dynamic Bus Line Adjustment System, to recommend new operating route for the existing bus line by building direct route between extracted hot bus stop pair. *DBLAS* mainly involves extracting hot bus stop pair based on passenger volume estimation, and planning optimal route between hot bus stop pair using taxi traces. Finally, we develop a demo system to demonstrate the effectiveness of *DBLAS*.

Keywords: bus route; passenger volume; hot bus stop pair

1 Introduction

During peak hours, a majority of bus lines are crowded due of limited capacity of existing urban public transport system. The intuitive solutions aim to shorten the dispatching interval between consecutive buses [3] and supplement with buses. Although it reduces residents' awaiting time for a bus, traveling time to the destination would probably become longer. This is due to that the condition of urban traffic block would become more aggravated because of the impacts of above measures. To improve residents' traveling experience, it necessitates an effective way to shorten traveling time based on the premise that the pressure of urban traffic remains unchanged.

There have been several researches on bus route adjustment over the years, but all of them attempt to build new operating route to cover more regions of the city [2] or enable ease of inhabitant trip at midnight [1]. Obviously that employing fixed operation scheme of bus routes cannot address our aforementioned issue, instead, designing various operation schemes for a same bus line is more advisable.

Observed that during morning rush hour most of residents generally take a trip by bus from the uptown and through the downtown to the workplace in the other side of city. For instance, through estimating passenger volume for each stop in a bus line using swiping card data at 8 a.m, we can see that passenger volumes of two bus stops are much greater than the others for a bus line. Additionally, such two bus stops do not locate on the downtown and

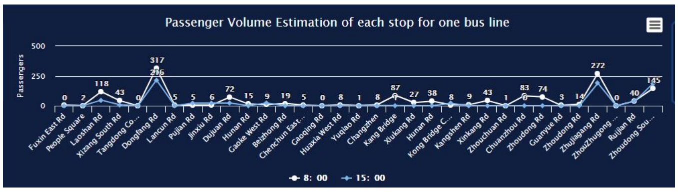


Fig. 1. Illustration of passenger volume estimation

distant from each other, as shown in figure 1. It indicates that a large percentage of residents who live at one end of city have to travel by bus through most congested roads to their destinations (e.g., workplace, public service agency, etc.) which locate on the other end of city. Identifying hot bus stop pair that is away from downtown, and building a direct route between them enables designing additional operating route for any bus line. In view of that, our route adjustment mechanism attend to extract hot bus stop pair through estimating passenger volume of each bus stop on any bus line, and obtain direct route between the extracted stops by planning less congested route using historical taxi traces.

2 Overview

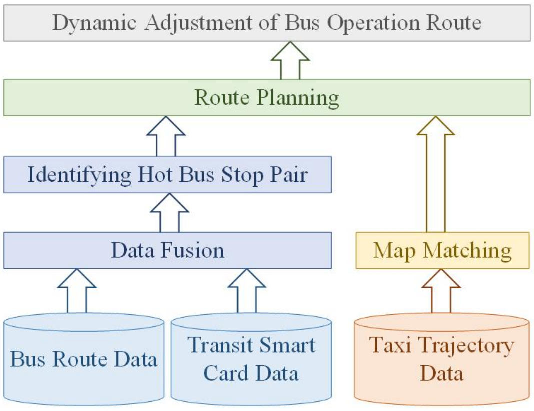


Fig. 2. DBLAS system

We propose a two-phase framework which consists of hot bus stop pair extraction and route planning between hot bus stops, as illustrated in Figure 2. At the first phase, we seek for the popular bus stops away from downtown by estimating passenger volume for each bus stop on one bus line, and then extract hot bus stop pair from them according to the specified threshold. At the second phase, we derive the direct route between hot bus stop pair through mining optimal route from the historical taxi traces. Noted that our route planning task is is premised on avoiding the worst of rush-hour congestion. Therefore, it is imperative to take traffic congestion situation into account when planning route.

Data Analysis. Three different data sets from Shanghai, China are used. (i) *Transit Smart Card Data* records the data of swiping the bus card; (ii)*Bus Route Data* collects the information of different bus routes; (iii) *Taxi Trajectory Data* contains the trajectories of taxis. The details of three data sets are shown in Figure 3. *Transit Smart Card Data* and *Bus Route Data* are used through data fusion to extract hot bus stop pair, and *Taxi Trajectory Data* is applied to derive the optimal route between hot bus stop pair. **Identification of Hot Bus**

Transit Smart Card Data		Bus Route Data		Taxi Trajectory Data	
Shanghai	3/2018	Shanghai	3/2018	Shanghai	4/2018
Size	9.95 GB	Size	57GB	Size	231.9GB
# of Records	8billion	# of Route	1265		
Format		Format		Format	
CardID	Date&Time	RouteID	BusID	TaxiID	Date&Time
RouteID		StationID	StartTime&EndTime	Speed	Longitude&Latitude

Fig. 3. Data sets

Stop Pair A bus route refers to an ordered set of bus stops, which is denoted as $P = \{p_1, p_2, ..., p_a, p_{a+1}, ..., p_{a+b}, ..., p_m\}$. Generally, bus stop setting strategy of a bus line adopts starting with uptown through downtown to uptown. Thus, bus stops located in downtown can be represented by $PD = \{p_a, p_{a+1}, ..., p_{a+b}\}$, $PD \subset P$. The core task of hot bus stop identification is to estimate passenger

Algorithm 1: Passenger Volume of Each Bus Stops Calculation

Input: $CSum$: the total number of transit smart card records;
$BusSet$: bus route data;
Output: $StopVolume$: a list of $StopVolume_{p_i}$;

1 $StopVolume \leftarrow \emptyset$;
2 $T_{Sum} = \sum_{i=1}^{m} T(p_i)$;
3 **foreach** *busstop* p_i ***in*** $BusSet$ **do**
4 $\quad StopVolume.add(\frac{T(p_i)}{T_{Sum}} \times CSum)$;
5 **return** $StopVolume$;

volume for each bus stop on a line. But it is difficult to accurately calculate passenger flow of each bus stop in case of only the total sum of all the bus stops is available. Given a time period (e.g., from 7:00 am to 10:00 am), we heuristically estimate passenger flow based on the ratio of the bus stoppage time of one bus stop to the total sum of bus stoppage time of all the bus stops, the detail process is illustrated in Algorithm 1. Let $T(p_i)$ denote the sum of stoppage time of all the buses that stop at one bus stop p_i, $i \in \{1, 2, ..., m\}$, $CSum$ denote the amount of transit smart card records within a specified time period. Through data fusion, the total sum of stoppage time of all the bus stops can be calculated by $T_{Sum} = \sum_{i=1}^{m} T(p_i)$ (lines 2-4), and hence passenger flow of bus stop p_i can be estimated by $StopVolume_{p_i} = \frac{T(p_i)}{T_{Sum}} \times CSum$ (Lines 6-7). Given a volume threshold τ, a bus stop $p_i \in P$ is defined as a hot bus stop, iff $p_i \notin PD$ and $StopVolume_{p_i} \geq \tau$. Accordingly, a hot bus stop pair is composed of two hot bus stops that are distant from each other.

Route Planning We propose a route planning method incorporated with road congestion situation. The speed of each road segment within a given time period, denoted as $S_j (j \in \{1, 2, ..., n\})$, can be obtained by average speed calculation of its matched historical taxi trajectory data after map matching. Each optimal route extracted between hot bus stop pair using algorithm [4] consists of a list of road segments, the length (denoted as $length(line_k)$) of which is obtained by the sum of all the traversed road segments' lengths. Combined with the derived speed of road segment, the traveling cost of a route can be calculated by the sum of the ratio of each traversed road segment's length len_j to S_j, i.e., $length(line_k) = \sum_{j=1}^{n} \frac{len_j}{S_j}$. The direct route between hot bus stop pair is derived by choosing the route with shortest traveling cost.

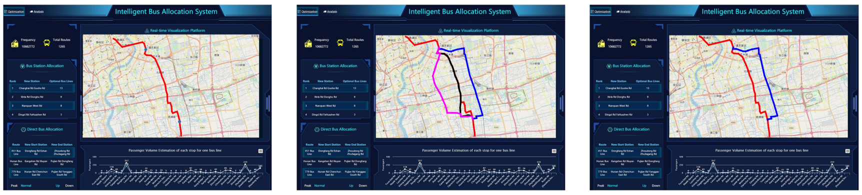
3 Demonstration

Via the system interface, different existing bus line can be selected for adjustment. Our demo system takes Shanghai No.451 bus line as an example.

Identification of Hot Bus Stop Pair.

A hot bus pair, as shown in Figure 4(a), <Zhoudong Road Zhujiagang Road→ Dongfang Lushan Road> is firstly derived by running *DBLAS* system, here the value of τ is empirically set to 200. We can see that there is a large software park near *Dongfang Lushan Road*, it means that a majority of residents who live near Zhoudong Road travel by No.451 bus line to there.

Route Planning. Then we plan the path between the two extracted hot bus stops using the method in [4], and obtain K (K is set to 4) paths with quite short lengths, as shown in Figure 4(b). Incorporated with road congested situation, we select a route with minimum traveling cost (depicted in blue line) as final result, who is near the original bus route (depicted in red line), as illustrated in Figure 4(c). Last but not least, we suggest thatnew derived route and original one should be used alternately within different time periods.



(a) A hot bus pair (b) Top 4 shortest routes (c) Final result

Fig. 4. DBLAS system demonstration

Acknowledgements The authors are very grateful to the editors and reviewers for their valuable comments and suggestions. This work is supported by NSFC(No.61702423, U1501252).

References

1. Chen, C., Zhang, D., Li, N., Zhou, Z.: B-planner: Planning bidirectional night bus routes using large-scale taxi GPS traces. Trans. Intelligent Transportation Systems 15(4), 1451–1465 (2014)
2. Chuah, S., Wu, H., Lu, Y., Yu, L., Bressan, S.: Bus routes design and optimization via taxi data analytics. In: CIKM. pp. 2417–2420 (2016)
3. Huang, Q., Jia, B., Jiang, R., Qiang, S.: Simulation-based optimization in a bidirectional A/B skip-stop bus service. IEEE Access 5, 15478–15489 (2017)
4. Luxen, D., Vetter, C.: Real-time routing with openstreetmap data. In: SIGSPATIAL. pp. 513–516. GIS '11 (2011)