Mobile phone GPS data in urban customized bus: Dynamic line design and emission reduction potentials analysis

Qing Yu a, b, Haoran Zhang b, *, Weifeng Li a, **, Xuan Song b, Dongyuan Yang a, Ryosuke Shibasaki b

a Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao’an Road, Shanghai, 201804, People’s Republic of China
b Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, Japan

A R T I C L E   I N F O

Article history:
Received 25 January 2020
Received in revised form 21 May 2020
Accepted 26 May 2020
Available online 4 July 2020

Handling editor: Dr Sandra Caeiro

Keywords:
Customized bus system
GPS data
Dynamic bus line planning
Emission reduction potentials

A B S T R A C T

Customized Bus is a new mode of Internet-supported public transportation. It is regarded as one of the major strategies to reduce the usage of private cars and mitigate greenhouse gas emissions from road traffic. For a customized bus system, a dynamic bus line planning system based on the demand can largely improve the performance and promote the public acceptance of customized bus service. This paper introduces a method to generate planning suggestions for bus lines and stops based on massive demand data. A link network is generated from the input to represent the sharing route of the demand. With community detection, the link network is segmented into communities with similar travel routes. By examining the core-peripheral structure and matching the core part of communities with the road network, the customized bus lines are generated. Boarding and alighting hotspots are identified as the suggestion for customized bus stops. The methodology is tested by using mobile phone data in Tokyo. With the input of one-day sample, the algorithm can generate the result in approximately 1 min and extract 29 bus lines. According to the shape and spatial location of the bus lines, three types of bus lines serving different travel patterns are classified: radiation type lines, ring-type lines, and suburban lines. Analyzing the emission reduction potential of the extracted bus lines manifests that bus line planning of the proposed method has the potential to relieve emission pressure on urban expressways and to reduce approximately 13% of road traffic emission.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Transportation is one of the major industries emitting greenhouse gas (GHG) which warms the earth and causes human-driven climate change (Montgomery, 2017; Watts et al., 2019). Today, the majority share of energy consumption and emissions occurs within cities or is related to the ongoing urbanization (Long et al., 2019). In 2016, GHG emissions from transportation were responsible for a quarter of total emissions, which has a close connection with urban sustainability (Kim, 2019; Song et al., 2020). Promoting an efficient public transportation system has been widely considered as a strategy to reduce emissions from road transportation in urban areas (Litman, 2012).

Customized Bus (CB) is a new Internet-supported public transport mode that is warmly applauded by the public in recent years (Liu and Ceder, 2015). Compared with the traditional public transit system, CB system is more efficient and better meets the demand of customers. Therefore, CB is regarded as an effective approach to attract more passengers, reduce the usage of private cars and mitigate emissions from road traffic. The performance of CB system largely depends on the system design. Currently, the CB systems in operation usually collect demand data from online surveys and manually plan CB lines, which is costly, inefficient and uncompetitive. A real-time dynamic CB line planning system based on passengers' demand can significantly improve the performance and promote public acceptance of CB system.

However, developing such a bus line planning system faces some new challenges: (1) CB systems need Internet-based platforms to collect passenger’s daily travel demand data efficiently.
The system should be capable of handling massive data input and generate results within short computation time. (3) The dynamic nature of passengers’ demand requires the algorithms of CB lines planning and stops distribution to be dynamic as well. Such algorithms should be integrated into the CB line planning system to generate CB networks based on real-time demand input or historical data.

This paper introduces a method to generate planning suggestions for CB lines and bus stop distribution based on massive demand data analysis. Using the car trajectories extracted from a mobile phone dataset as input, a case study was conducted in Tokyo and generates 29 CB lines. The rest of the paper is organized as follows: Section 2 reviews related works and provides a brief introduction of a dynamic CB line planning system; section 3 presents the specific problems related to the CB line planning system; Section 4 details the methodology of the CB line formation and hotspot analysis of CB stop distributions; Section 5 presents the results of the case study in Tokyo; Section 6 proposes conclusions and the contribution of this study.

2. Related works

2.1. The concept of CB line

As an emerging sustainable travel mode in the public transit system, CB has drawn great attention recently. The existing study of customized bus systems mainly focuses on the following aspects: 1. analysis of CB concept (Liu and Ceder, 2015), 2. CB network design (Han et al., 2019), 3. CB management strategies (Chuanyu et al., 2017) and 4. passenger preference on CB (Cao and Wang, 2016).

The concept of CB service was introduced as a kind of subscription bus services several decades ago. Research has been mainly focused on the concept of subscription bus services, guidelines for subscription bus network, the operational planning and the pricing strategy, etc (Bautz, 1975; Chang and Schonfeld, 1991; Kirby and Bhatt, 1975; McCall Jr, 1977; McKnight and Paaswell, 1985; Potts et al., 2010). In recent years, Qingdao launched the first CB system in August 2013 in China (Liu and Ceder, 2015). Liu et al. introduced the concept of the new customized bus and provide a systematic examination and analysis of the development and current state of CB practices in China (Liu and Ceder, 2015).

Lyu et al. proposed a typical design of CB lines (Fig. 1). Along the CB bus line, there are two types of bus stops: 1. Grouped bus stops: multiple bus stops are arranged in a small region to assemble travelers. 2. Intermediate bus stops: few intermediate bus stops are arranged along the bus route to guarantee the efficiency of CB line. In this design, the CB bus can provide a more efficient and direct service than a traditional bus.

As a successful and typical practice of CB mode, E-drive is a CB company committed to exploring and implementing CB service in Shanghai (E-DRIVE, 2019). E-drive offers several types of CB service including long-term CB service called micro railway, park area CB service, temporary leases CB, and special customized line, etc. The operation procedures of E-drive is Internet-based. Customers can submit their demand including the information of their origins, destinations and departure times via online platforms such as web sites and mobile phone apps. CB company will match the demand and plan bus lines for customers with similar travel demand. In their micro railway service, e-drive divides the area of Shanghai into 11 sub-areas and offers 45 CB lines to connect sub-areas. Customers can check the existing CB lines and stations, bus timetables, real-time bus GPS location, submit orders and reserve bus seats via online platforms.

2.2. CB network design

Research of CB network design mainly focuses on designing CB lines direction and stop location by using optimization models. These models usually based on the following steps (Li et al., 2018; Liu and Ceder, 2015): demand extraction; CB lines and stops deployment; further optimization on scheduling, dispatching and controlling, etc. In detail design of CB systems, research has well-developed dozens of models that can consider multiple factors, including dynamic routing and timetabling, congestion detouring, etc (Ma et al., 2017b).

Several studies have handled city-scale datasets to extract potential CB demand and establish models for CB stops and route design. Ma et al. proposed a hierarchical clustering-based methodological framework based on origin-destination (OD) region division, route OD region pairing, and route selection (Ma et al., 2017a). Instead of considering the OD as a connection between city regions, this methodology handles OD statically and separately, which failed to offer suggestions for a demand-oriented CB route design. Qiu et al. use an improved density-based spatial clustering of application with noise (DN) algorithm to discovered potential CB demand and generate candidate location of stops using smart card data (Qiu et al., 2018). In this study, they mainly focus on identifying, estimating and clustering potential CB passengers from bus commuters in a large urban area, without generating suggestions for CB route design.

Lyu et al. systematically introduced the concept of CB network design and proposed a planning framework called CB-Planner (Lyu et al., 2019). The differences between the method presented in this study and CB-Planner are as follows: (1) CB-Planner initially deploys bus stops and then connect the stops with CB lines. However, as the backbone of cities, urban transportation systems are, in large measure, shaping the spatial structure of the city (Anas et al., 1998; Gong et al., 2017; Zhong et al., 2014). From the city planning aspect, as the focus has been shifted to designing demand from serving demand, to design the bus line and decide how the bus line connects the areas into integration becomes increasingly important. Therefore, in the method proposed, bus line direction is decided before deploying bus stops. (2) Similar to most existing CB planning models, CB-Planner generates CB lines and CB stops with the exact location. However, stops should be located in the meaningful and stoppable places. Generating results completely depending on the optimization model may not be suitable in the real-world situations and difficult to implement. The aim of the methodology introduced here is to generate suggestions on line direction and a range of regions to set bus stops with possibilities for further detailed design. The bus line planner can further design and adjust the detailed route choice and bus stop location according to the real-world situation.

In summary, the existing CB design models have several week points: 1. Most of the existing methodologies handle the origin point and the destination point of passengers’ demand statically and separately, without a comprehensive consideration of the
A dynamic CB line planning system is key to CB service. Fig. 2 shows an overview of the system. First of all, passengers will submit their demand via the online platform. At set intervals, the demand will be uploaded into the cloud server and aggregated into origin (O) and destination (D) as the input of the dynamic CB line planning system. Then, the system will search for the best route choice for each OD pairs. Based on the routes, the system will plan CB lines in the city in order to meet the demand. The stops of CB lines will then be deployed so that the CB system can publish CB riding information to passengers and inform them of locations of stops and departure time.

In this system, the most difficult part is the planning of CB lines. The aim is to generate the route of CB lines based on the massive volume of travel routes. The system has to decide the number of CB lines to be deployed and the route of each CB line as well. Obviously, the more CB lines deployed as possible, the more demands will be served, but at the same time, the more of the CB lines will be uneconomical with less demand for each CB line (Lyu et al., 2019). So the solution to this problem involves the balance between two objectives: 1. Few lines, less convinence and high profits for each line 2. Large number of lines, great convinence, but low profits for each line.

Planning a bus line is a very complicated task in real-world situations. Many factors will affect the chosen of the bus line and the efficiency of the bus transit system, including the policies, the planning process, and the existing technology (Ibarra-Rojas et al., 2015). The resulting bus lines are usually a compromise of interest from three parties: government, bus operation company and passenger (Nayan and Wang, 2017).

The results produced by the methodology in this study will answer the following questions: 1. The CB line will serve which part of the transportation demand and connect which part of the city? Given that, how to design the route of CB lines? 2. Which range of areas along the CB lines is suitable for the deployment of bus stops?

### 2.3. Examining mobility patterns and extracting sharing potential from big data

With the development of information technology, large-scale spatio-temporal data emerge and has been widely used in examining human mobility patterns, such as include mobile phone data, public transit smart card data, taxi GPS data, etc. Many researchers dedicate to efficiently express the massive and large-scale people’s movement and uncover regional travel patterns and regularities at an urban scale (Qiu et al., 2018). In recent years, the emergence of the sharing economy has also attracted a great deal of attention. Research has provided evidence that sharing transportation services can significantly decrease the traffic congestion (Li et al., 2016), facilitate the first-and last-mile public transit connections (Shahheen and Chan, 2016) and improve social, economic, and environmental sustainability in a city (Wu and Zhi, 2016). Base on spatio-temporal big data, predict mobility sharing potential at an urban scale will be significant support for building shared transportation services beneficial to citizens, businesses and society (Yu et al., 2020).

### 3. Problem description

A dynamic CB line planning system is key to CB service. Fig. 2 shows an overview of the system. First of all, passengers will submit their demand via the online platform. At set intervals, the demand will be uploaded into the cloud server and aggregated into origin (O) and destination (D) as the input of the dynamic CB line planning system. Then, the system will search for the best route choice for each OD pairs. Based on the routes, the system will plan CB lines in the city in order to meet the demand. The stops of CB lines will then be deployed so that the CB system can publish CB riding information to passengers and inform them of locations of stops and departure time.

In this system, the most difficult part is the planning of CB lines. The aim is to generate the route of CB lines based on the massive volume of travel routes. The system has to decide the number of CB lines to be deployed and the route of each CB line as well. Obviously, the more CB lines deployed as possible, the more demands will be served, but at the same time, the more of the CB lines will be uneconomical with less demand for each CB line (Lyu et al., 2019). So the solution to this problem involves the balance between two objectives: 1. Few lines, less convinence and high profits for each line 2. Large number of lines, great convinence, but low profits for each line.
4.2. Construction of link network

The basic concept of CB line design is to discover the clusters of trajectories sharing their routes. And the important routes in each cluster indicate the route of CB line. To accomplish this target, instead of matching and clustering the trajectories, the method proposed in this study generates a link network to describe the sharing routes from the trajectory data set and to segment the vertices of the network into network communities which indicate the travel demand clusters.

First of all, the spatial resolution is decided for matching the sharing route of trajectories. In our previous study, we have proved that when the resolution of spatial grids is set smaller than 500 m, the activity intensity of users will be largely affected by the temporal resolution. On the other hand, when the spatial grids are larger than or equal to 500 m, the influence of the temporal resolution will be largely eliminated (WeiFeng, 2018). After considering the user’s activity intensity and the characteristics of the dataset, the spatial resolution here is set as 500 m mesh grids. Considering a set of trajectory $T = \{\text{Traj}_1, \text{Traj}_2, ..., \text{Traj}_{\text{numtra}}\}$, each trajectory is denoted as $\text{Traj}_i = \{(p_{i1},l_{i1}), (p_{i2},l_{i2}), ..., (p_{ilen},l_{ilen})\}$. Where $\text{numtra}$ is the number of trajectories, $ilen$ is the length of $\text{Traj}_i$. $loc_i$ is the index of the mesh grid representing the location of point $p_i$. Each trajectory can also be transformed into the set of links $\text{Traj}_i = \{(l_{p1}, l_{p2}), (l_{p2}, l_{p3}), ..., (l_{pilen-1}, l_{pilen})\}$. As the links between mesh grids being considered as vertices in a network, each trajectory indicates a set of connections between vertices. The dataset of trajectories can generate a weighted undirected link network. The detail of link network construction is shown in Algorithm 1.

The advantage of constructing a link network from trajectory is that the most important routes which the trajectories shared the link network construction is shown in Algorithm 1. After detecting link communities from the link network, to determine which vertices are part of a densely connected core and community to which vertex is assigned, the function $\delta(c_i, c_j)$ is defined as Eq. (2)

\[
\delta(c_i, c_j) = \begin{cases} 
1 & c_i = c_j \\
0 & c_i \neq c_j 
\end{cases}
\]

In order to maximize the modularity efficiently, the fast unfolding algorithm repeatedly iterates modularity optimization and community aggregation to obtain the maximum of global modularity. For detailed information on the algorithm, please refer to (Blondel et al., 2008). This study adopted the fast unfolding algorithm provided in igraph python package (Csardi and Nepusz, 2006) for this study.

4.4. Detecting core-peripheral structure in communities

After detecting link communities from the link network, to determine which vertices are part of a densely connected core and...
which are part of a sparsely connected periphery in link communities allow us to extract the important sharing routes in the travel demand clusters. Here, the Rombach’s algorithm is used to find continuous core-periphery structure in each community. The algorithm is a quantitative method to investigate the core-periphery structure, which computes a continuous value called ‘coreness’ for each vertex in the network. The objective of the algorithm proposed ‘coreness’ measurement maximizes the core quality of the network. The core quality is formulated as Eq. (3)

$$\begin{align*}
R &= \sum_{ij} A_{ij} C_i C_j \\
\end{align*}$$

(3)

where $A_{ij}$ is the weight of the edge between vertices $i$ and $j$, and it equals to 0 if vertices $i$ and $j$ are not connected and $C_i$ denotes the local coreness of the vertex $i$, which is given by Eq. (4)

$$C_i = \begin{cases} 
\frac{i(1 - \alpha)}{2\beta}, & i \in \{1, \ldots, \beta N\} \\
\frac{(i - \beta)(1 - \alpha)}{2(N - \beta)} + 1 + \frac{\alpha}{2}, & i \in \{\beta N + 1, \ldots, N\}
\end{cases}$$

(4)

where $N$ is the total number of vertices, parameter $\alpha$ sets the size of the score jump between the highest-scoring periphery vertex and the lowest scoring core vertex and the parameter $\beta$ sets the size of the core. The objective of the algorithm is to find a shuffle of vertex that yields $C_i$ for each vertex which maximizes the core quality $R$ of the network. To mitigate the computational cost, a label switching algorithm is implemented in the algorithm. Here, the parameters are set to be $\alpha = 1$ and $\beta = 0.8$, in order to classify each vertex to either the core or the periphery, and extract the vertices with the top 20% coreness as core vertices in a community. For detailed information on the algorithm, please refer to literature (Rombach et al., 2017).

### 4.5. Extracting CB line route and identifying potential demand

The densely connected core vertices of link communities are the set of links connecting mesh grids. For the reason that the link network is constructed by car trajectories, the core links are mostly distributed along trunk roads in the city. Geographic Information System (GIS) based approach is applied here to extract CB line route from the set of links. The extracting method contains three steps: 1. Select the communities with the line shape and capable to extract
CB line. 2. Merge the core links within a tolerance distance and simplified into a single line. In this study, the tolerance distance is set to be 500 m. 3. Match the simplified line into the road network as bus line route.

After extracting CB lines, a rule is set to identify the potential demand for each line: 1. Generate a buffer area with a tolerance distance for each CB line. Here, the tolerance distance is set as the same value in CB line extracting 2. For each trajectory, the proportion of the trajectory shared with the CB line can be calculated as Eq. (5)

\[ p = \frac{l_{\text{shared}}}{l_{\text{total}}} \]  

where \( l_{\text{shared}} \) is the length of the sub trajectory inside the buffer area, and \( l_{\text{total}} \) is the total length of the trajectory. The trajectory with the \( p \) over 80% can be regarded as the potential travel demand of the CB line, indicating that over 80% part of the trajectory is sharing the same route with the CB line.

Passenger flow per kilometer is the indicator widely adopted to evaluate the operation of public transit lines. In the Guidelines for optimization of bus (electric bus) lines in Shanghai developed by the Shanghai Traffic Committee, this index is adopted as the criterion of determining the necessity of opening a new transit line and the optimization of existing bus lines (Committee, 2016). Here, the index of potential demand trajectories per kilometer is adopted as the index of measuring the potential operation performance and benefits of each CB line, which is calculated as Eq. (6)

\[ O_i = \frac{n_i}{d_i} \]  

where \( n_i \) is the number of potential travel demand (trajectories here) of CB line \( i \), and \( d_i \) is the length of the CB line. A higher value of \( O_i \) indicates that with the same level of operation cost, the CB line can attract more potential travel demand with better operational benefits.

4.6. Identifying boarding and alighting hotspot for CB stop

For determining the locations most suitable for setting the bus stops for each CB line, it is significantly helpful to identify the spatial hotspot clusters of potential boarding and alighting location. After extracting the potential travel demand of CB lines, the first point and the last point of the sub trajectory inside the buffer area can be regarded as the potential boarding and alighting location of this trajectory.

Here, the local Moran’s I index is introduced to compare the sample value with its neighbors and identify the hotspot of boarding and alighting locations (Anselin, 1995). The high positive value of this index demonstrates that the demand for boarding and alighting demand in the grid is similar to the demand for boarding and alighting in the adjacent grid. Two types of hotspots can be identified when both considering this index and the value of the sample: the “regional hotspots” and the “individual hotspots”. Regional hotspots indicate that the sample is a hotspot surrounded by high-value samples; On the other hand, Individual hotspots indicate that the sample is a hotspot surrounded by low-value samples.

In the scenario of CB line boarding and alighting hotspot identification, the “regional hotspots” and “individual hotspots” are both the hotspots with significantly higher demand for boarding and alighting demand along the CB lines. The two types of hotspots have different demand patterns so that they should be considered to set different types of CB stations.

4.7. Emission model for potential emissions reduction

In this paper, the COPERT (COmputer Programme to calculate the Emissions from Road Transport) model is adopted for measuring the fuel consumption (FC) and three types of emission including carbon monoxide (CO), nitrogen oxides (NOx) and hydrocarbon (HC) for each car trajectories. Using COPERT model, Burón et al. analyzed the emissions of several local, global, and fuel-related vehicular pollutants, and discussed further the reasons behind (Burón et al., 2004). Bellasio et al. estimated the road traffic emission inventory and evaluated the contribution of different vehicle categories to the emissions (Bellasio et al., 2007). This model is also successfully used in measuring emission using car GPS data (Sui et al., 2019). In this study, the COPERT model is used for estimating the emission reduction potential of CB lines. It is assumed that the CB system can
Fig. 7. Example of Link communities and CB line extracted.

Fig. 8. CB lines planning suggestions from link communities. (a) The direction of 29 CB lines extracted (b) Hotspot identified for bus stop.
has been processed for confidentiality is used, the anonymization processing is performed by NTT DOCOMO INC. For further detail of the dataset, please refer to Supplementary material B.

After identifying the travel modes from the dataset, the car trajectory data is sampled as the simulation of dynamic demand input of the algorithm. Fig. 5 shows the distribution of the car trajectories. Implementing the methodology on the platform with Intel i7-8650U CPU and 16 GB RAM. For a one-day sample data taken from the dataset (32,647 car trajectories in total), the algorithm proposed takes 18.55 s to construct the link network, 9.36 s to community detection and 43.57 s to detect the core-peripheral structure of the network. The total computation time to generate the result is approximately 1 min.

5.2. Result of community detection and core-peripheral structure in communities

From the car trajectory dataset, a network is generated with 189,829 vertices and 2,336,738 edges. The community detection algorithm produces a result of community segmentation with the modularity of 0.73, indicating that there is a community structure in the link network. In the result of community segmentation, there are 10,313 communities in total, 851 of communities are with over 10 links, and only 152 communities are with over 100 links. The complementary cumulative distribution function (CCDF) of the number of links in link communities is shown in Fig. 6. The CCDF curve decreases quickly in the small number of links, indicating that a relatively small number of communities contain a significantly large number of links and play an essential role in the link network. Examine the core-peripheral structure of the large communities, most of the core links of link communities are in a line shape and suitable to extract bus line. Fig. 7 shows two examples of the core-peripheral structure of link communities and the bus line extracted.

5.3. Result of CB line route and bus stop hotspots

From the core-peripheral structure of 152 link communities with over 100 links, 29 CB lines are extracted in the area of Tokyo. The average length of the CB lines is 18.7 km, with the longest of 38.1 km and the shortest of 11.6 km. The number of CB lines is smaller than the number of link communities, and the reason is that the CB lines extracted from some of the small communities are much shorter, most of them are in the same direction as the larger communities. The travel demand from these small link communities can be satisfied by the CB lines extracted from large communities if the location of bus stops are properly arranged. The core-peripheral structure of link communities that extracts the 29 CB lines is shown in Supplementary material C.

After extracting CB lines, the potential demand is identified for each line. The CB lines are ranked in descending order from 1 to 29 according to the number of potential demand (Fig. 8(a)). For better understand the function of CB lines and further provide operation and management suggestion of future application for the CB project, the CB lines are classified into three types and named based on the shape of bus lines and the place they connected:

- Radiation type lines: This type of CB line include line 2, 3, 4, 5, 7, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 25 and 29. These CB lines are extending from city center to suburban area in all directions and offering an efficient transportation method for passengers from some specific suburban residential area to enter the city center or sub-center rapidly. These CB lines are mostly in the straight-line shape as the shortest path from the origin to the destination to ensure the efficiency, which
may serve as the commuting line for people living in suburban area and working in the city center.

- Ring-type lines: This type of CB line include line 1, 6, 8, 9, 11, 22, 26 and 28. This type of CB line is on the edge of the central urban district. These CB lines will mostly serve the business and tourism demand by connecting sub-centers of the city.

- Suburban lines: This type of CB line include line 10, 18, 24 and 27. These CB lines connect multiple suburban centers and serve multiple travel purposes.

For CB lines in different type, they should have different operational and scheduling strategies. These three types of CB lines are reasonable and in line with the position of CB lines as a supplement in the urban transportation system. Notice that the presenting CB lines network does not connect multiple Central Business Districts (CBDs) in one line. The reason is that the travel demand from one CBD to another can be well served by urban subway lines.

Fig. 8 (b) shows the regional hotspots and individual hotspots identified from boarding and alighting of potential demand from each CB line. Regional hotspots indicate the cluster of high demand locations along the CB line, which can be the location for grouped bus stops. Grouped stops in a range of areas can assemble passengers and reduce their walk distance to the bus stops; Individual hotspots indicate a high demand location in a low demand neighborhood. Intermediate bus stops can be arranged in these hotspots. Arranging these two types of stops according to the hotspots can ensure the satisfaction of demand without devastating the efficiency and directness of the CB line.

5.4. Evaluation of travel demand for CB lines

To sum up the number of potential demand, there are 290,465 trajectories in total, which can be potentially replaced by the CB travel mode (20.6% of the total number of all car trajectories and 13.1% of the total mileage).

Fig. 9 shows the operation benefit of CB lines. The operation benefit of each CB line is around 400 to 600 trajectories per kilometer, indicating that with proper line detail design and operation management, CB lines will have a similar level of operational performance. Among all the CB lines, line 2 is the one with the highest operation benefit, connecting Kawasaki city directly to the city center.

Fig. 10 shows the hourly change of the CB demand, which suggests the bus schedule arrangement and pricing strategy. In the hourly demand, the morning peak is around 6:00 to 8:00, and the evening peak is around 16:00 to 18:00. However, the hourly demand distributed differently among CB lines. One of the demand patterns is the morning peak and evening peak can be clearly identified, including CB line 1, 2, 3, 4, 5, 6, 10, 13, 14, 15, 16, 20, 21 and 28. These CB lines are in a tide traffic demand pattern with a high proportion of commuting trips. On the contrary, the morning and evening peak hours of CB lines 7, 8, 9, 11, 12, 17, 18, 19, 22, 23, 24, 25,
26, 27 and 29 are not clear, with high demand during all day. This pattern indicates that travel demand of these lines has less repeatability.

Fig. 11 shows the box plot of the number of potential demands for each bus line with the comparison of weekdays and weekends & holidays. The number of potential demand fluctuates within a small range, indicating that the CB lines will have sustainable demand. Comparing the weekdays with weekends & holidays, in general, the potential demand for CB lines is more significant and more stable on weekdays. Travel demand on weekdays is more stable with high repeatability.

5.5. Result of emission reduction potential

Based on the speed and travel length of potential trips, the FC and emission are calculated. The potential FC and emission reduction can be calculated by summing the emissions produced by the car trajectories covered by each CB line. Fig. 12 shows the summing of the proportion of the potential emission reduction of each CB line to the total car emission in the whole city. The result shows that by implementing the entire 29 CB lines, there will be respectively 13.6%, 13.4%, 13.0% and 12.8% potential reduction of NOx, FC, HC and CO.

Fig. 13 shows the distribution of potential emission reduction, in which the spatial heat map shows the relative values ranging from 0 to 1, compared to the maximum volume, averaged by four types of emissions. As is shown in Fig. 13, a large part of potential emissions handled by CB lines is distributed on the periphery area of the city center, especially the southwest part of the city. A large part of emissions is concentrated at the Central Business Districts (CBDs), such as Setagaya, Shibuya, Shinagawa, and Toshima. The reason of this distribution pattern is that the three types of CB lines developed by the methodology can serve the travel demand from suburban areas to CBD areas in the city center, such travel highly depends on the urban expressway. Therefore, CB lines can potentially reduce emission pressure on urban expressways by replacing private car travel.

6. Conclusions and future prospects

As a new mode of Internet-based transportation in the public transportation system, the CB system has great potential to replace private cars and reduce road traffic emissions in urban areas. However, the efficiency of such a demand-oriented system hinges upon public acceptance and operation performance. In light of this, this paper proposed a methodology to generate planning suggestions for CB lines and stops based on massive demand data analysis. The algorithm of which, at high computing speed, is capable of being integrated into a dynamic CB planning system. The proposed method involves the following steps:

The car trajectories are extracted from mobile phone data as the input, potential CB travel demand. Based on the demand input, a link network is generated to represent the generalized, common routes of passengers. The link network is segmented into
communities with similar travel routes by community detection and the core-peripheral structure of each community is examined. By extracting the core parts of the link network communities and matching them with the road network, CB lines are formed. Potential demand for CB lines is identified and boarding and alighting hotspots are extracted as the suggestion for CB stops. The methodology is tested by using “Konzatsu-Tokei(R)” mobile phone data in Tokyo. Using one-day sample, the algorithm can generate the result in approximately 1 min and extract 29 bus lines. According to the shape and spatial location of the CB lines, three types of CB lines serving different travel patterns are classified, including radiation type lines, ring-type lines and suburban lines. By analyzing the emission reduction potential of the extracted CB lines, the proposed method has the potential to relieve emission pressure on urban expressways and to reduce approximately 13% of road traffic emission.

Practical implications of the present study:

- The method proposed can automatically generate the routes of bus lines and stop locations suggestions based on massive demand data input, which is more demand-oriented and efficient than the existing traditional CB line planning method.

- The algorithm is capable of processing massive demand data and generates results in time, which has great potential of being integrated into a dynamic CB line planning system.

- By analyzing the case study in Tokyo, the CB lines generated by the proposed method have the potential to relieve emission pressure on urban expressways and reduce approximately 13% of road traffic emission.

In future studies, the proposed method has several potential research and application directions:

Firstly, there are many factors affecting FC and car emissions, such as vehicle type, load, slope of roads, number of passengers, vehicle manufacturing, driver behavior including driving habits, braking frequency, parking frequency, etc. However, from the GPS data, it is difficult to obtain relevant information and conditions for such accurate calculation. In future studies, more detailed data can be applied to conduct accurate emission analysis.

Secondly, a more reasonable model for emission reduction potential estimation is to consider the emission removed by cars and add the emissions imposed by using buses. Urban buses also produce high emissions that have energy and environmental impacts on urban sustainability. Studies have focused on proposing methodologies for estimating the FC and emissions from urban buses (Jia et al., 2018; López-Martínez et al., 2017; Rajaeief et al., 2019; Ribau and Silva, 2017; Wang et al., 2020; Yu et al., 2016). However, the methodology of this paper is to generate planning suggestions for bus lines and stops, which do not include the arrangement and scheduling of buses. Therefore, the emissions of buses in actual operation have not been considered in this study. Bus and crew dispatching model can be developed based on the bus lines generated from the method presented in this study, on which the further detailed emission calculation model can be established. In the real-world application, CB system with dynamic bus line planning system, dispatching system, and real-time controlling system can be developed and integrated.

It will take some time for a rising, new transportation service system to cultivate devoted customers and optimize the operation mode. The method we put forward is designed to help concerning institutions and enterprises to find potential passengers’ demand and to grasp the future of urban transportation. As a potential direction of future studies, examining the operational performance and finding a better solution for customized public transportation service still requires integrating the research results with passenger willingness, operation practices, and many other factors.

Credit authorship contribution statement

Qing Yu: Conceptualization, Methodology, Software, Writing - original draft, Data curation, Visualization. Haoran Zhang: Writing - review & editing, Conceptualization. Weifeng Li: Writing - review & editing. Xuan Song: Data curation. Dongyuan Yang: Supervision, Funding acquisition. Ryoosuke Shibasaki: Supervision, Resources, Project administration.

Declaration of competing interest

There are no conflicts of interest to declare.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 71734004).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2020.122471.

References
