Chin. Geogra. Sci. 2021 Vol. 31 No. 1 pp. 27−40
https://doi.org/10.1007/s11769-021-1174-z

Springer Science Press
www.springerlink.com/content/1002-0063

Delineation of an Urban Community Life Circle Based on a Machine-Learning Estimation of Spatiotemporal Behavioral Demand

LI Chunjiang, XIA Wanqu, CHAI Yanwei

(1. *College of Urban and Environmental Sciences*, *Peking University*, *Beijing* 100871, *China*)

Abstract: Delineating life circles is an essential prerequisite for urban community life circle planning. Recent studies combined the environmental contexts with residents' global positioning system (GPS) data to delineate the life circles. This method, however, is constrained by GPS data, and it can only be applied in the GPS surveyed communities. To address this limitation, this study developed a generalizable delineation method without the constraint of behavioral data. According to previous research, the community life circle consists of the walking-accessible range and internal structure. The core task to develop the generalizable method was to estimate the spatiotemporal behavioral demand for each plot of land to acquire the internal structure of the life circle, as the range can be delineated primarily based on environmental data. Therefore, behavioral demand estimation models were established through logistic regression and machine learning techniques, including decision trees and ensemble learning. The model with the lowest error rate was chosen as the final estimation model for each type of land. Finally, we used a community without GPS data as an example to demonstrate the effectiveness of the estimation models and delineation method. This article extends the existing literature by introducing spatiotemporal behavioral demand estimation models, which learn the relationships between environmental contexts, population composition and the existing delineated results based on GPS data to delineate the internal structure of the community life circle without employing behavioral data. Furthermore, the proposed method and delineation results also contributes to facilities adjustments and location selections in life circle planning, people-oriented transformation in urban planning, and activity space estimation of the population in evaluating and improving the urban policies.

Keywords: community life circle; spatiotemporal behavioral demand; demand estimation model; decision tree; ensemble learning

Citation: LI Chunjiang, XIA Wanqu, CHAI Yanwei, 2021. Delineation of an Urban Community Life Circle Based on a Machine-Learning Estimation of Spatiotemporal Behavioral Demand. *Chinese Geographical Science*, 31(1): 27−40. https://doi.org/10.1007/s11769-021-1174-z

1 Introduction

The 'National New-type Urbanization Plan' of China necessitates the transformation of urban planning from place-oriented to people-oriented, from emphasizing quantity to balancing both the quantity and quality of urban development, and from focusing on economic production to recognizing residents' needs for better lives (The Central Committee of the Communist Party of China and the State Council of China, 2014). As a key direction for the transformation of urban planning in China, urban life circle planning considers the daily life of the residents as a planning object and can unify physical spatial planning, behavioral planning and social planning. The goal of urban life circle planning is to realize equal and precise allocation of public service facilities, satisfy the increasingly variable needs of residents, and promote the realization of bottom-up and parti-

Received date: 2020-04-07; accepted date: 2020-08-23
Foundation item: Under the auspices of the National Natural Science Foundation of China (No. 41571144)
Corresponding author: CHAI Yanwei. E-mail: chyw@pku.edu.cn

cipatory planning (Wu, 2015; Sun and Chai, 2017).

An urban life circle is a geographical space where residents usually conduct their daily lives, and it consists of the necessary spatiotemporal resources (Chai, 2014; Liu and Chai, 2015). Specifically, an urban life circle exists at the following five spatial scales: community, housing cluster, commute, metropolitan and urban agglomeration life circles (Chai et al., 2015). The community life circle is the entry point for the transformation of urban planning and has recently received considerable attention from practitioners and researchers in urban planning. To evaluate the facility supply, traditional residential area planning adopts methods such as 'index per thousand people' and 'service radius', which can not fulfill the actual facility related needs of residents while improving the residential satisfaction and sense of community (Xu and Ye, 2010; Sun and Chai, 2017). In this context, developing novel planning theories and methods from the perspective of the community life circle can help overcome the shortcomings of traditional residential area planning (Yu, 2019). The newly released standard for urban residential area planning and design (GB 50180−2018) introduces the concept and ideas of the community life circle, and takes the 5-min, 10-min and 15-min community life circles as the main planning objects (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2018). Major cities in China, such as Beijing, Shanghai, Jinan and Changsha, have already attempted to plan the 15-min community life circle to improve the quality of life and residential satisfaction (Municipal Bureau of Planning and Natural Resources of Shanghai, 2016; Municipal Bureau of Planning and Natural Resources of Jinan, 2019).

However, many aspects pertaining to the planning of community life circles are still under discussion, including the concept definition, scope delineation, function affiliation, and the development of planning methods and implementation frameworks (Chai and Li, 2019). Among these aspects, formulating the strategies to identify and delineate the scope of a life circle is a prerequisite for further planning. The existing delineation methods are facility-based and pay insufficient attention to the community microenvironments, composition and needs of residents, and this configuration does not conform with people-oriented community life circle planning. Most of the research and planning practices are based only on the indicators to delineate the scope of a life circle, such as the areas accessible within a certain duration by walking, population size and land area, and are coordinated with the administrative boundaries (Municipal Bureau of Planning and Natural Resources of Shanghai, 2016; Municipal Bureau of Planning and Natural Resources of Jinan, 2019; Guo et al., 2019; Han et al., 2019). Other scholars delineated the community life circle considering the completeness and density of the public service facilities (Cui et al., 2016; Xiao et al., 2018). However, such delineation results lack flexibility and can not satisfy the new requirements to adapt to local contexts and respond to the various needs of residents in community life circle planning (Yu, 2019).

Theoretically, the community life circle is equivalent to the aggregated result of the activity space near and within the community of all the residents (Chai and Li, 2019). Activity space is defined as 'the local area within which people move or travel in their daily activities' (Wang et al., 2018). Compared with the traditional facility-based buffer method, by delineating the community life circles from the activity space, the differences in the socio-economic status, walking ability and spatiotemporal demand of residents in different communities can be fully considered (Rainham et al., 2010; Chai et al., 2015). However, the existing activity space delineation methods involve certain limitations when used to delineate community life circles. Conventionally, the activity space has been measured by standard deviational ellipses, GPS trajectory buffers, minimum convex polygenes and kernel density surfaces (Perchoux et al., 2013; Sharp et al., 2015). Such methods may be biased by the scale of the community life circle, and certain parameters may often be defined arbitrarily. Sun et al. (2016) considered the area in which the residents conducted non-work activities outside their homes as the activity space and community life circle; however, the result was limited by the sample size of the behavior surveys. Several scholars identified the activity space using location-based service (LBS) data, such as cellular signaling data and check-in data (Ahas et al., 2015; Zhen et al., 2017). However, the scale of such data is excessively small to effectively delineate the life circle of a community. In addition, the existing methods only take into account the actual behavior and neglect the role of the environmental contexts (Wang et al., 2018). The identified activity space is thus notably constrained by the behavioral data and can not precisely describe the potential activity space.

In response to this imperfection, Chai et al. (2019) recently proposed a delineation method for community life circles, based on the 'context-based crystal-growth (CCG) activity space' method, which can sufficiently integrate the behavioral information and environmental contexts. In this approach, an accessibility-weighted plane is firstly defined based on the environmental contexts, then the walking accessible range is delineated with a high precision by using the CCG method, and finally the internal structure within the range is identified using GPS data (Wang et al., 2018; Chai et al., 2019). The walking accessible range is the maximum area accessible by walking within a certain period of time under the constraints of the environmental contexts, which are reflected by the accessibility-weighted plane, and those of the walking ability of residents of different ages and different abilities. Due to the low acquisition cost of the environmental data and residential demographic composition, as they can be derived from the existing data sets, the range delineation method can be generalized to other communities in a relatively easy manner. The internal structure of the community life circle represents the spatiotemporal demand for the different facilities within the walking-accessible range. Specifically, this structure reflects the core concept of the community life circles by focusing on the various needs of the residents, in contrast to the concept pertaining to the traditional residential areas. The identification of the internal structure relies considerably on behavioral data, especially GPS data. However, the cost of GPS surveys is extremely high, and thus, this method can not be easily applied to other communities.

To realize community life circle planning, a delineation method that can integrate the behavioral information with environmental contexts and can be easily generalized is required. The CCG activity space method can facilitate such integration, but it is not generalizable. The key reason is that the identification of the internal structure depends on the behavioral data. Therefore, this study was aimed at developing a method based on which the internal structure could be delineated through established relationships among the behavioral demands and certain easily obtained variables, instead of new surveyed behavioral data. Machine learning techniques were applied to establish the estimation models for the spatiotemporal behavioral demand for facilities within the walking-accessible range. By combining the CCG method with the estimation models, an urban community life circle delineation method that is generalizable and can integrate humans and the environment was developed.

2 Method

2.1 Framework

The framework of the community life circle delineation method based on the spatiotemporal behavioral demand estimation is shown in Fig. 1. This method can be applied to communities for which GPS data are or are not available.

First, the life circles were delineated according to the existing CCG activity space method in communities with the GPS data. This method included two stages: 1) delineating the 15-min walking-accessible range by using the CCG method based on the environmental contexts and population composition, and 2) delineating the internal structure by conducting a kernel density analysis based mainly on the behavioral data.

Second, the spatiotemporal demand estimation models were constructed by analyzing the relationships between the behavioral demand reflected by the internal structure, environmental contexts and population composition. The key challenge in ensuring the applicability of the delineation method is the high cost of obtaining the behavioral data. The internal structure, the identification of which depends significantly on the behavioral data, reflects the spatiotemporal behavioral demand for different areas within the walking-accessible range. Therefore, estimating the behavioral demand and thereby delineating the internal structure by using low-cost data, such as environmental contexts and population composition, can help delineate the life circles in communities without behavioral data.

To this end, machine learning techniques were applied to construct the demand estimation models. Machine learning has received increasing attention from researchers in urban and behavioral studies, because of its high estimation performance and the relaxation of the assumptions in traditional regressions. In transportation behavior research, machine learning techniques have

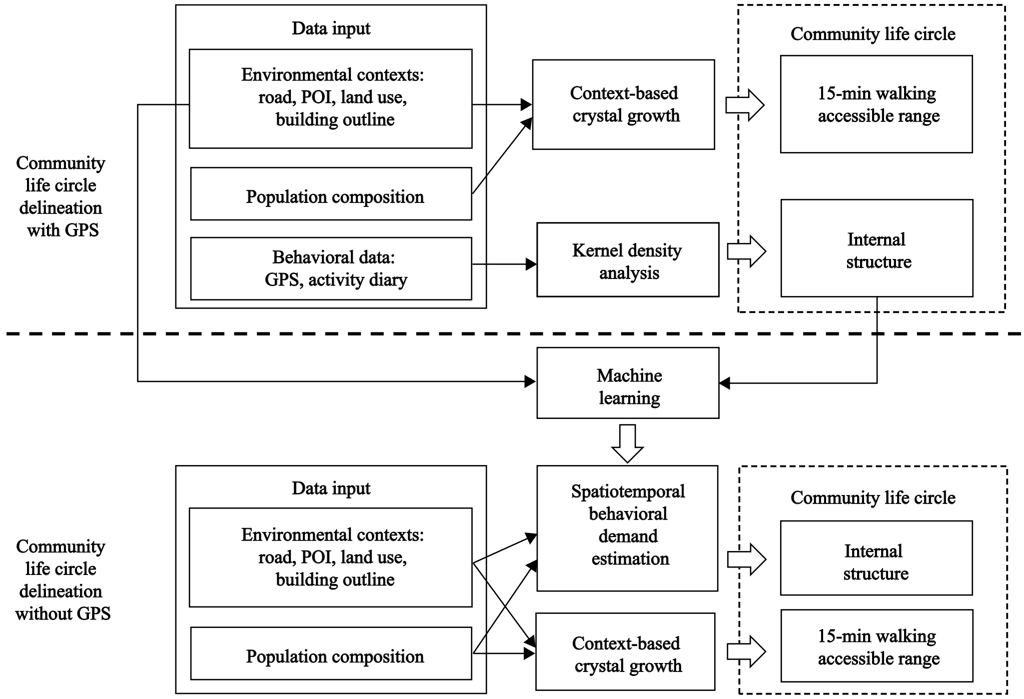

Fig. 1 Framework of the community life circle delineation method based on spatiotemporal behavioral demand estimation

already been applied for different purposes. For example, decision trees have been used to model the individual spatiotemporal behavior and to derive the behavior decision rules from activity travel data (Arentze et al., 2000; Arentze and Timmermans, 2004; Sammour and Vanhoof, 2018). Regional travel demand has also been modeled and predicted using data mining techniques (Ghasri et al., 2017). Other applications include the modeling of travel mode choices (Tang et al., 2015; Hagenauer and Helbich, 2017; Wang and Ross, 2018), modeling of walking route choices (Tribby et al., 2017), and inferring trip purposes by combining smart card data and activity travel diary data (Alsger et al., 2018). However, machine learning is still underrepresented in the research of spatiotemporal behavioral demand modeling. This study can contribute to the existing research in this regard.

Finally, the life circles in the communities without GPS data were delineated by combining the CCG method and the spatiotemporal behavioral demand estimation models obtained from machine learning. The CCG method was used to delineate the 15-min walking-accessible range based on the environmental contexts and population composition, and the demand estimation model was applied to identify the internal structure within this range. This delineation method required data that could be acquired at a low cost, and thus, the generalizability was ensured.

2.2 Model construction

The key task in this work was the construction of the spatiotemporal behavioral demand estimation models according to the framework of the delineation method. The construction procedure is defined in Fig. 2. Four stages were implemented to obtain the estimation models: identifying the spatiotemporal behavioral demand of the residents, selecting the explanatory variables, applying the machine learning techniques and choosing the final models with the lowest error rates.

First, the identified internal structures of the communities with GPS data were treated as the residents' spatiotemporal behavioral demand for different plots of land within the 15-min walking-accessible range. The plot of land was used as the basic analysis unit because it is also the basic object in urban planning. Considering that residents may have different demand patterns for different land use types, demand estimation models for five types of land use, namely, public service, commercial service, green land, residential land and other types were developed, according to the code for the classification of land use for urban and rural planning of Beijing (DB 11/996−2013). The other types consisted of industrial land, utility land and other land types that were unrelated to the function of the community life circle. The number of GPS points per capita on each plot served as the spatiotemporal behavioral demand for the plot (Table 1). In general, the absolute number of GPS points is meaningless because the value is affected by the sample size and duration of behavioral surveys. Therefore, the demand for each plot was defined as high or low according to the median value of this number. The plots with a number of GPS points greater than or equal to and fewer than the median corresponded to 'high demand' and 'low demand', respectively.

Second, explanatory variables were selected to estimate the spatiotemporal behavioral demand, including the land characteristics, community population composition and built environment. Two aspects were considered when selecting these variables. First, the acquisition cost of the variables was required to be low; for example, an existing database was available for the variables, which guaranteed the generalizability of the delineation method. Second, the explanatory variables were required to be correlated with the behavioral demand. In previous studies, the researchers considered the distance, demographic variables such as age and

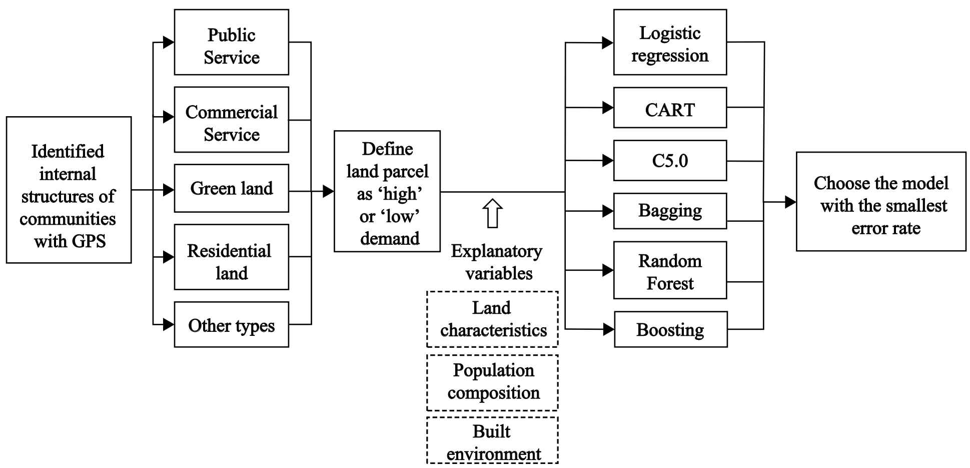

Fig. 2 Process of constructing the spatiotemporal behavioral demand estimation models

Table 1 The variables for construction of the spatiotemporal behavioral demand estimation models

Variables	Description
Dependent variable	
Spatiotemporal behavioral demand	The total number of the non-work non-travel GPS points outside the home on the plot divided by the sample size of the community, classified as high or low demand according to the median value
Explanatory variables	
Distance	Distance to the community center of each plot
Area	Area of each plot
Age	Proportion of different age groups, including children (0–14), young adults (15–29), the middle-aged group (30–49) and the elderly group (≥50)
Education	Proportion of residents with education levels below high school (*ref.* level) and high school and above
Hukou	Proportion of local residents with *hukou* and migrants without *hukou* (*ref.* level)
Public facility density	Number of public facilities divided by the plot area
Commercial facility density	Number of commercial facilities divided by the plot area
Diversity	Simpson diversity index of Point of interest (POIs) (Comer and Greene, 2015)
Transit accessibility	Distance to the nearest bus stop
Else variable	
Land use type	Types of land use, including public service, commercial service, green land, residential land and other types

education, as well as the built environment when constructing the behavioral models (Tang et al., 2015; Ghasri et al., 2017; Hagenauer and Helbich, 2017). Considering these aspects, nine explanatory variables belonging to three categories were selected (Table 1). Among these variables, the distance and area of each plot represented the basic land characteristics. A larger distance of the plot from the center of the community and a smaller plot area corresponded to a lower spatiotemporal behavioral demand for the plot. The age, education level and *hukou* status structures represented the different needs of the different residents for the land. In general, the household registration (*hukou*) status has been noted to be a key socio-economic variable that also influences the travel behavior (Li and Liu, 2016; Zhang et al., 2018). The introduction of the sociodemographic composition into the estimation models can ensure that the delineated life circle can represent the community characteristics. In terms of the built environment, the density of the public and commercial service facilities, diversity of points of interest (POIs) and transit accessibility were considered. These variables were noted to be closely related with the spatiotemporal behavior (Ewing and Cervero, 2010; Hagenauer and Helbich, 2017).

Third, different machine learning techniques as well as logistic regression were applied to model the behavioral demand for different types of plots. Logistic regression, which is the most widely used analytical model for behavioral decisions, served as the basic model to demonstrate the strengths of the machine learning techniques. In this work, decision tree and tree-based ensemble learning were applied as the machine learning techniques. Tree-based learning methods have been applied for many purposes, for instance, to analyze the mode choice, trip purpose, travel destination and spatiotemporal behavioral decisions (Arentze and Timmermans, 2004; Hagenauer and Helbich, 2017; Ghasri et al., 2017). These methods generally exhibit a higher performance and efficiency than those of logit models and artificial neural networks (Xie et al., 2003, Tribby et al., 2017).

Decision trees utilize a tree-like structure for data classification. Starting with the root, each node recursively splits the data by features, and the leaves represent the classes. ID3 and classification and regression trees (CART) are the most widely used decision tree induction algorithms. ID3 uses the entropy measure to choose the attribute at each node. C4.5 improves upon the ID3 algorithm (Quinlan, 1993), and C5.0 further improves upon the C4.5. Therefore, in this research, C5.0 was employed. The CART algorithm, which is based on the Gini index, measures the purity of a response distribution and evaluates the splits (Breiman et al., 1984). Although decision trees can effectively manage nonlinear relationships, a single tree is sensitive to noise and tends to overfit (Hagenauer and Helbich, 2017). Tree-based ensemble techniques combine many decision trees to obtain a higher predictive performance than that of any single classifier. Bagging trains classifiers in a parallel manner by using bootstrap samples. Each classifier has an equal weight, and the majority vote determines the class assignment in the prediction (Breiman, 1996). The random forest (RF) algorithm is similar to bagging. Although the random forest also trains classifiers using bootstrap samples, the nodes of the trees are determined by a random subset of variables (Breiman, 2001). Boosting is different from the bagging and random forest techniques in that it trains classifiers successively, with a new classifier established to improve the incorrect classifications in the preceding classifiers. The prediction is based on weighted voting (Freund and Schapire, 1997). In this work, the adaptive boosting (AdaBoost) technique was applied, which is the most commonly implemented type of boosting.

All the techniques were implemented for each type of land use in the R programming environment (R Core Team, 2019). The relevant packages for this research were ‘raprt’ (Therneau et al., 2019), ‘C50’ (Kuhn et al., 2020), ‘adabag’ (Alfaro et al., 2013) and ‘randomForest’ (Liaw and Wiener, 2002). The performance of each model was estimated using 10-fold cross-validation, to reduce the bias in selecting the training and testing subsets (Kohavi, 1995). The parameters of the number of trees in the ensemble and the number of variables randomly sampled at each split in the random forest were defined by trial and error, and the parameters of the models that produced the lowest error rate were chosen as the final parameters. The other parameters were assigned default values.

Finally, for each type of land use, the model with the lowest error rate was chosen as the final estimation model. For the final models, the variable importance (VI) was determined through the algorithms to examine the effectiveness of the selected explanatory variables. The mean decrease in the Gini index of each variable was used as the Ⅵ measure (Ghasri et al., 2017). Although more effective approaches are available to define the Ⅵ, the average reduction in the Gini index was used in this research because the Ⅵ among the different ensemble learning models is comparable.

3 Data

3.1 Data source

The data for this research were primarily derived from the 'daily activity and travel survey of Beijing residents', which was conducted by the Behaviorlal Research Group of Peking University in the Qinghe sub-district (*jiedao*), Haidian District, Beijing in 2012. Overall, the continuous 7-d GPS data and 7-d activity diary data for 242 valid residents in 15 communities (Fig. 3) and community population composition information were included. In addition, the environmental context data for the survey communities, including roads, POIs, land use and building outlines were employed. The data for the 14 communities were applied to construct the spatiotemporal behavioral demand estimation models. The remaining Dangdai Chengshi Jiayuan (DDJY) community was used as an example to delineate the community life circle based on the models. The predicted result was compared with the actual result to verify the effectiveness of the delineation method. The DDJY community was chosen as the example as it is a representative community of the Qinghe sub-district, in which most of the communities developed at approximately 2000. Moreover, this community consists of different types of residents and is located in the center of the district.

The land plots were considered as the basic analysis units and divided into five types: public service, commercial service, green land, residential land and other types. Although the life circles of different communities may share the same plots, the plots in different life circles were treated as distinct units because certain explanatory variables were different in these regions. The final numbers of the analysis units were as follows: 72 public service plots, 75 commercial service plots, 101 green land plots, 158 residential land plots and 126 other types of plots.

3.2 Data preprocessing

Data preprocessing was performed to obtain the vari-

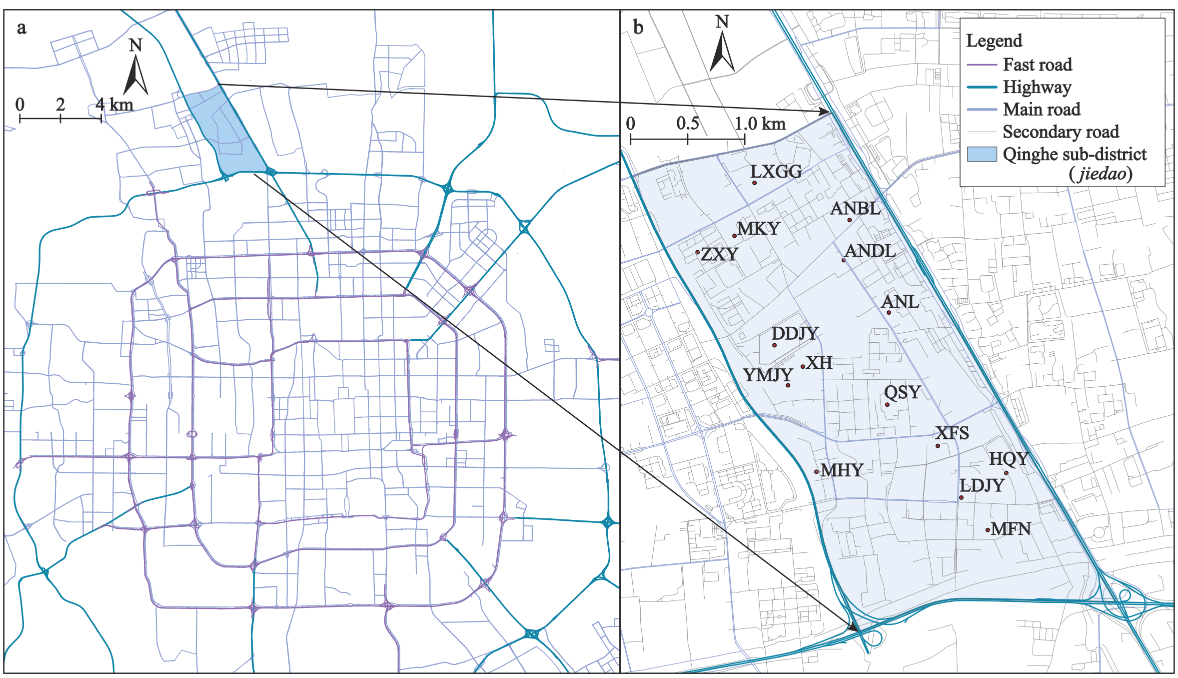

Fig. 3 Survery area: a) location of Qinghe Sub-district, Beijing; b) locations of the surveyed communities in Qinghe sub-district, Beijing. ANBL: Anning Beilu; ANDL: Anning Donglu; ANL: Anningli; DDJY: Dangdai Chenshi Jiayuan; HQY: Haiqingyuan; LDJY: Lidu Jiayuan; LXGG: Lingxiu Guigu; MFN: Maofangnan; MHY: Meiheyuan; MKY: Mingkeyuan; QSY: Qingshangyuan; XFS: Xuefushu Jiayuan; YMJY: Yimei Jiayuan; ZXY: Zhixueyuan

ables required to construct the estimation models.

First, the life circles of the 14 communities (without DDJY) were delineated by applying the CCG activity space method (Chai et al., 2019). The population composition was used to calculate the walking ability of different groups of residents. By combining the walking ability with the environmental context data, including the roads, land use and building outlines, 15-min walking-accessible ranges were delineated by using the crystal-growth algorithm. Therefore, the ranges were spatially different among different communities. The GPS data and activity diaries were further applied to obtain internal structures within the accessible range.

Second, we obtained the behavioral demand for the plots in each community life circle. The accessible plots were identified by overlaying the plots of land with the ranges of the community life circles. For a certain community, the spatiotemporal behavioral demand for each accessible plot within the life circle was represented by the internal structure, which was defined as the number of non-work non-travel GPS points outside the home per capita (Fig. 4). The plot was classified as having a high or low demand by considering the median value when the estimation models were constructed. In addition, other variables were calculated, including the distance to the center of the community of each plot, area of each plot, and built environment variables. Along with the community population composition, these variables were treated as the explanatory variables required to construct the estimation models (Table 1).

4 Results

4.1 Estimation models

The spatiotemporal behavioral demand estimation models for five types of land use were constructed using six methods, namely, logistic regression, CART, C5.0, bagging, random forest and boosting. The error rate was estimated using the mean misclassification rates in ten folds. The rate in the logistic regression model was extremely higher compared to those in the decision tree models, which demonstrated the effectiveness of machine learning (Fig. 5). The rates in the single tree models were usually higher than those in the ensemble learning models, especially for public service and other types, which indicated that the behavioral demand was only weakly associated with the explanatory variables and that the ensemble learning techniques could improve the performance. For each type of land, the model with the lowest error rate was chosen as the final estimation model. Therefore, the random forest models were chosen as the final models for the public service, commercial service, residential land and other types. The boosting model was chosen as the final model for green land. The error rates in the final models were 33.3%, 28.6%, 37.6%, 41.1% and 36.7%. The performance of the final spatiotemporal behavioral demand estimation models was comparable to that of the previous demand models, the accuracy of which varied from 17% to 72% depending on the model and its definition of accuracy (Arentze and Timmermans, 2004; Ghasri et al., 2017).

The variable importance indicates the impact of the explanatory variables on the prediction of the dependent variable. Fig. 6 shows the VIs in the final estimation models of different types of land use. The values were scaled to 1 to compare the different models by ensuring that the largest VI in each model was 1 (Ghasri et al., 2017). The most important variable for all the models was the plot distance, indicating the law of the distance

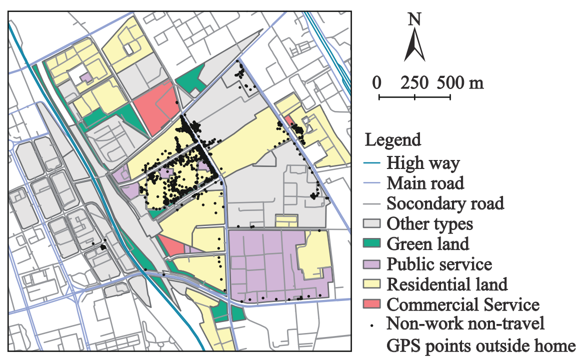

Fig. 4 Data preprocessing result for a certain community (DDJY) in Beijing

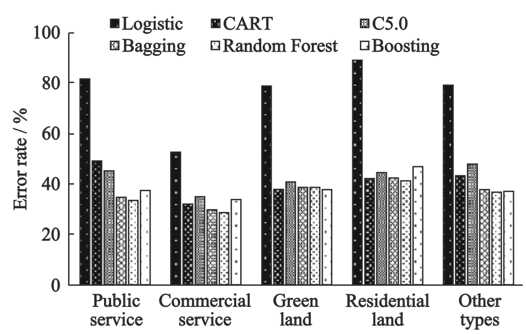

Fig. 5 Error rates of the estimation models for different types of land corresponding to different methods

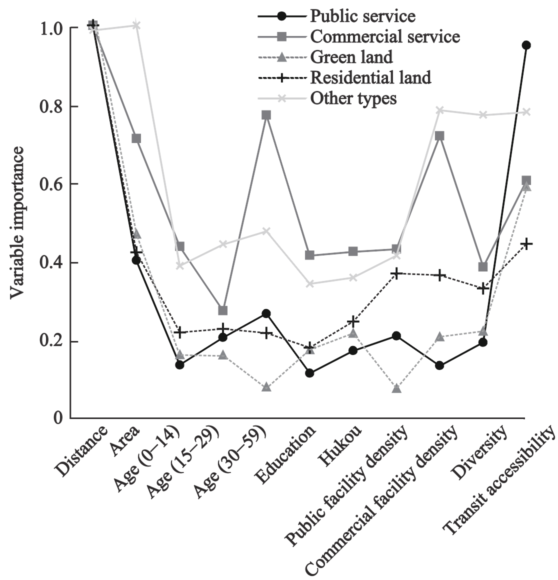

Fig. 6 Variable importance (VIs) in the estimation models of different types of land use

decay. The importance ranking for the other variables was more complex and influenced by different types of land use. Except for the distance, the important variables for the public service model were the transit accessibility and plot area. Surprisingly, the importance of the public facility density was relatively small. The results indicated that the demands for the public service were most considerably influenced by the plot characteristics and transit accessibility, instead of the concentration of the facilities. The plot area, age structure between 30–59, commercial facility density and transit accessibility exhibited a high importance in the commercial service model. This observation suggested that the middle age group may have a higher demand for commercial service, especially for plots with a larger area, higher density and higher accessibility. Moreover, the remaining variables, except for that pertaining to young adults (aged 14–29) exhibited a relatively high importance compared to that in the other models. The important variables for the green land model were similar to those of the public service model because green infrastructure is also regarded as public service. The area and transit accessibility were important variables for the residential land model. Moreover, the facility density and diversity also exhibited a relatively high importance. The results indicated that the plot characteristics and built environment of the residential land were critical to attract residents. Finally, for the model of the other land types, consisting of various types of land use, most of the variables exhibited a relatively high importance. The importance ranking was as follows: plot characteristics, built environment, age structure, *hukou* and education.

Overall, the plot characteristics represented the most important variable, and the distance was more important than the area. The results confirmed the findings of Hagenauer and Helbich (2017), who found that the trip distance is the most important variable. The built environment was the second most important variable in most models, and among the built environment variables, the transit accessibility was the most important. The second most important variable was the commercial facility density, and the diversity and public facility density ranked third. Most of the population composition variables corresponded to the lowest importance. Nevertheless, the importance of the middle age group (aged 30–59) was relatively higher than that of the other groups, especially in the models for the commercial service and other types. These results suggested that the environmental contexts likely exert a notable influence on the spatiotemporal behavioral demand, while the impact of the demographic variables is modest. In addition, the results were consistent with those of the existing research, which found the land use characteristics were the most important factors in predicting the travel demand, and age was relatively important compared to the education and income (Ghasri et al., 2017; Hagenauer and Helbich, 2017).

4.2 Framework validation

Subsequently, we considered the DDJY community as an example to demonstrate the mechanism of the estimation models to delineate the community life circle without GPS data based on the framework shown in Fig. 1. First, the 15-min walking-accessible range was delineated using the CCG method by employing the environmental contexts and population composition (Fig. 7A). Second, the accessible plots were identified by overlaying the range with the land use. The independent variables used to estimate the spatiotemporal demand were obtained by extracting the types of land use; evaluating the plot distance, plot area and built environment; and linking the population composition structure of the DDJY. Third, the demand of the residents for each plot (high or low) was predicted by applying the estimation models. Finally, the internal structure

Fig. 7 Delineated life circle in the Dangdai Chengshi Jiayuan (DDJY) community: (A) 15-min walking-accessible range; delineation result based on the (B) estimated demand and (C) actual demand

was identified by overlaying the estimation results and accessible range (Fig. 7B).

In the comparison of the delineated community life circle based on the estimated and actual demands (Fig. 7B and Fig. 7C), the prediction accuracy was at an acceptable level of 63.0% in the plots with behavioral data. In these plots, the 14 actual high-demand plots were predicted as 12 high-demand and 2 low-demand plots, and the 13 actual low-demand plots were predicted as 8 high-demand and 5 low-demand plots. The residential land and other types of plots were most often misclassified, specifically, 2 of the misclassified high-demand plots corresponded to residential land, and 4 of the misclassified low-demand plots corresponded to 2 other types and residential land plots each. Because the other types involved different types of land use, the low accuracy in estimating the corresponding demand was expected. For residential land, the misclassification likely occurred owing to the popular gates and surrounding walls of the communities in the Chinese urban residential areas (Douglass et al., 2012). In general, gated communities make it challenging to estimate the demand for residential land, and this aspect is consistent with the highest error rate pertaining to its estimation model. However, community life circle planning focuses mainly on matching the demand and supply of public and commercial service. Estimating the demand for residential land and other unrelated land use types is less important than estimating the public service, commercial service and green land. Moreover, if the plots of residential land and other types are eliminated, an accuracy of 86.7% can be attained. Therefore, the performance of the estimation models is acceptable, and the models can be applied for community life circle planning.

The estimation models also yielded the prediction results for the plots without behavioral data. Forty-four plots without GPS points were predicted as 23 high-demand plots and 21 low-demand plots. However, there were a variety of reasons that the plots had no behavioral data, such as survey samples, survey times and facilities distributions, and thus, the efficiency of the estimation models in these areas could not be evaluated.

5 Discussion

This research proposed a community life circle delineation method based on a machine-learning estimation of the spatiotemporal behavioral demand. In addition to delineating the life circles without GPS data, the method can be applied for other purposes in urban planning.

First, the estimated behavioral demand can cover areas in no behavior data are available within the 15-min walking-accessible range, which provides a reference for realizing the facility adjustments and location selection in future community life circle planning. The areas for which the residents indicate a high demand (high-demand plots) but do not visit at present are valuable objects in community life circle planning. In this context, the constraints on walking should be reduced, and the walking environment should be improved in these areas. Moreover, new facilities should be placed in these areas because these regions already have the potential to attract residents. By removing the walking constraints, improving the walking environment and adding new facilities, community life circle planning can provide more opportunities for conducting daily life activities, thereby improving the quality of life.

Second, the estimated behavioral demand for the different types of land use in different community life circles can help overcome the limitations of the inflexibility of the traditional residential area planning. The index per thousand people method is commonly used in traditional planning; however, this approach fails to fit the various demands of residents. The estimated behavioral demands and delineated community life circles are unique for different communities, because they take into account the community population composition, environmental contexts and resident behavior. Therefore, planning new facilities and adjusting the old facilities based on the estimated results can satisfy certain needs of the residents and improve their quality of life.

Third, the behavioral demand estimation models can be used not only in community life circle delineation but also to estimate the activity space in large areas. By conducting behavior surveys in certain sampled communities and establishing the relationships among the activity space and demographic and environmental variables, we can estimate the activity space of the complete population as well as that for the different groups of people. Estimating the spatiotemporal behavioral demand and activity space of the general and different groups, especially children, aged and disabled individuals, is the first step to evaluate and improve the urban policies and urban planning process.

Although the demand estimation models are generalizable and useful for delineating community life circles and can help evolve the current urban planning methods. Several issues must be addressed in future research.

First, the transferability of the estimation models should be validated. In this study, the generalizability of the models was tested on a holdout sample. Because the holdout sample was derived from the same survey, the test was focused more on the internal validity than the external validity (Arentze and Timmermans, 2004). Because the survey area only represents the suburban areas of the megacities in China, it remains unclear how well the models would perform in another study area. The transferability of a model to a larger area has recently received attention in the disaggregated travel demand system of models (Ghasri et al., 2017). Although the transferability of the estimation models remains questionable, the framework could be transferred to other study areas. Estimation models should be constructed in other contexts based on the local behavioral surveys and environmental data, such as the city center or outer suburban areas, to improve the external validity of the model.

Second, the performance of the models should be improved. The small sample size negatively influenced the prediction accuracy in this study. Therefore, larger datasets must be used in future research to train the estimation models to improve their performance and robustness. Furthermore, more variables having higher correlations with the behavioral demand along with a low-acquisition cost should be included. At present, big data, such as cellular signaling data, provide opportunities for acquiring certain important factors at a low cost and with a large sample size, such as the proportion of the commuters and family structures (Ahas et al., 2015). With an expanded sample size, another aspect can be considered to improve the performance. Specifically, the types of land use can be further differentiated according to the residents' daily activity demand, such as schools, hospitals and gymnasiums, because different activities exhibit different relationships with the built environment and socio-economic status. Moreover, the goal of community life circle planning is to match the supply and demand of facilities (Sun and Chai, 2017). Future research can also model the demand for different types of facilities, which can provide more direct reference for the actual planning.

Third, the spatiotemporal behavioral demand should be identified with a higher precision. Future research can delineate life circles for different groups of people, considering their different walking abilities and behavioral patterns. This aspect can help identify and solve the problems faced by the disadvantaged groups when using community facilities. Furthermore, the integration of the information and communication technologies (ICTs) in everyday life has led to a considerable increase in online activities, which involve certain interactions with offline activities and alter the relationships among the behavior and environment contexts (Thulin et al., 2020; Xi et al., 2020a, b). E-activities may also challenge space-time constraints, the fixed relations between e-activity and physical activity space, and thus, have an influence on physical space (Kwan, 2007; Schwanen and Kwan, 2008; Wang et al., 2015; Loo and Wang, 2018). Future community life circle research and planning should consider the e-activities and their influence as well. Finally, both the spatial and temporal characteristics of the behavior demand should be considered. The temporal characteristics of the behavioral demand for a certain facility include not only the amount of time that the residents require this facility, which can be represented by the number of GPS points, but also at the times that the residents use it, which was not considered in this work.

6 Conclusions

The development of the delineation method for urban community life circles is one of the most popular research directions in the academic study and practice of urban planning. However, the delineation methods in most existing studies need behavioral data, of which the acquisition cost is so high that it is difficult to generalize the methods. This research improved the existing methods by introducing spatiotemporal behavioral demand estimation models that learn the relationships among the environmental contexts, population composition and existing delineated results based on GPS data to delineate the internal structure of the community life circle without behavioral data. Furthermore, the life circle of a certain community was delineated as an example by applying the estimation models that employed only the easily obtained environmental contexts and community population composition. The delineation result based on the estimated demand corresponded to an acceptable predictive accuracy compared with the result based on the actual demand, which indicated the effectiveness of the estimation models.

References

Ahas R, Aasa A, Yuan Y et al., 2015. Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, 29(11): 2017–2039. doi: 10.1080/13658816.2015.1063151

Alfaro E, Gamez M, García N, 2013. Adabag: an R package for classification with boosting and bagging. *Journal of Statistical Software*, 54(2): 1–35. doi: 10.18637/jss.v054.i02

Alsger A, Tavassoli A, Mesbah M et al., 2018. Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies*, 87: 123–137. doi: 10.1016/j.trc.2017.12.016

Arentze T A, Hofman F, Van Mourik H et al., 2000. Using decision tree induction systems for modeling space-time behavior. *Geographical Analysis*, 32(4): 330–350. doi: 10.1111/j.1538-4632.2000.tb00431.x

Arentze T A, Timmermans H J P, 2004. A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7): 613–633. doi: 10.1016/j.trb.2002.10.001

Breiman L, Friedman J, Stone C J et al., 1984. *Classification and Regression Trees*. London: Chapman and Hall/CRC.

Breiman L, 1996. Bagging predictors. *Machine Learning*, 24(2): 123–140. doi: 10.1023/A:1018054314350

Breiman L, 2001. Random forests. *Machine Learning*, 45(1): 5–32. doi: 10.1023/A:1010933404324

Chai Y W, 2014. From socialist *danwei* to new *danwei*: a daily-life-based framework for sustainable development in urban China. *Asian Geographer*, 31(2): 183–190. doi: 10.1080/10225706.2014.942948

Chai Yanwei, Zhang Xue, Sun Daosheng, 2015. A study on life circle planning based on space time behavioural analysis: a case study of Beijing. *Urban Planning Forum*, (3): 61–69. (in Chinese)

Chai Yanwei, Li Chunjiang, 2019. Urban life cycle planning: from research to practice. *City Planning Review*, 43(5): 9–16, 60. (in Chinese)

Chai Yanwei, Li Chunjiang, Xia Wanqu et al., 2019. Study on the delineation model of urban community life circle: based on qinghe district in Haidian district, Beijing. *Urban Development Studies*, 26(9): 1–8, 68. (in Chinese)

Comer D, Greene J S, 2015. The development and application of a land use diversity index for Oklahoma City, OK. *Applied Geography*, 60: 46–57. doi: 10.1016/j.apgeog.2015.02.015

Cui Zhenzhen, Huang Xiaochun, He Lianna et al., 2016. Study on urban life convenience index based on POI data. *Geomatics World*, 23(3): 27–33. (in Chinese)

Douglass M, Wissink B, Van Kempen R, 2012. Enclave urbanism in China: consequences and interpretations. *Urban Geography*, 33(2): 167–182. doi: 10.2747/0272-3638.33.2.167

Ewing R, Cervero R, 2010. Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76(3): 265–294. doi: 10.1080/01944361003766766

Freund Y, Schapire R E, 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1): 119–139. doi: 10.1006/jcss.1997.1504

Ghasri M, Rashidi T H, Waller S T, 2017. Developing a disaggregate travel demand system of models using data mining techniques. *Transportation Research Part A: Policy and Practice*, 105: 138–153. doi: 10.1016/j.tra.2017.08.020

Guo Rong, Li Yuan, Huang Mengshi, 2019. Research on optimization strategy of walking network in 15-minute community life circle of Harbin. *Planners*, 35(4): 18–24. (in Chinese)

Hagenauer J, Helbich M, 2017. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78: 273–282. doi: 10.1016/j.

eswa.2017.01.057

Han Zenglin, Li Yuan, Liu Tianbao et al., 2019. Spatial differentiation of public service facilities' configuration in community life circle: a case study of Shahekou district in Dalian city. *Progress in Geography*, 38(11): 1701–1711. (in Chinese). doi: 10.18306/dlkxjz.2019.11.006

Kohavi R, 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. San Francisco: ACM, 1137–1145.

Kuhn M, Weston S, Coulter N et al., 2020. *C50: C5.0 Decision Trees and Rule-Based Models*. R package version 0.1.3.1. Available at: https://cran.r-project.org/web/packages/C50/C50.pdf

Kwan M P, 2007. Mobile communications, social networks, and urban travel: hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer*, 59(4): 434–446. doi: 10.1111/j.1467-9272.2007.00633.x

Li S M, Liu Y, 2016. The jobs-housing relationship and commuting in Guangzhou, China: Hukou and dual structure. *Journal of Transport Geography*, 54: 286–294. doi: 10.1016/j.jtrangeo.2016.06.014

Liaw A, Wiener M, 2002. Classification and regression by randomForest. *R News*, 2(3): 18–22.

Liu T B, Chai Y W, 2015. Daily life circle reconstruction: a scheme for sustainable development in urban China. *Habitat International*, 50: 250–260. doi: 10.1016/j.habitatint.2015.08.038

Loo B P Y, Wang B, 2018. Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation*, 45(2): 365–384. doi: 10.1007/s11116-017-9792-0

Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2018. Standard for urban residential area planning and design. http://www.mohurd.gov.cn/wjfb/201811/t20181130_238590.html. Cited 15 January 2020. (in Chinese)

Municipal Bureau of Planning and Natural Resources of Shanghai, 2016. Shanghai planning guidance of 15-minute community life circle. http://ghzyj.sh.gov.cn/zcfg/ghss/201609/t20160902_693401.html. Cited 15 January 2020. (in Chinese)

Municipal Bureau of Planning and Natural Resources of Ji'nan, 2019. Jinnan planning guidance of 15-minute community. http://jnup.jinan.gov.cn/art/2019/1/31/art_10231_2824958.html. Cited 15 January 2020. (in Chinese)

Perchoux C, Chaix B, Cummins S et al., 2013. Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. *Health & Place*, 21: 86–93. doi: 10.1016/j.healthplace.2013.01.005

Quinlan J R, 1993. *C4.5: programs for Machine Learning*. San Mateo: Morgan Kaufmann Publishers.

R Core Team, 2019. *R: a Language and Environment for Statistical Computing*. Vienna, Austria: R for Statistical Compu ting.

Rainham D, McDowell I, Krewski D et al., 2010. Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. *Social Science & Medicine*, 70(5): 668–676. doi: 10.1016/j.socscimed.2009.10.035

Sammour G, Vanhoof K, 2018. A validation measure for computational scheduler activity-based transportation models based on sequence alignment methods. *Transportation Planning and Technology*, 41(7): 736–751. doi: 10.1080/03081060.2018.1504183

Schwanen T, Kwan M P, 2008. The internet, mobile phone and space-time constraints. *Geoforum*, 39(3): 1362–1377. doi: 10.1016/j.geoforum.2007.11.005

Sharp G, Denney J T, Kimbro R T, 2015. Multiple contexts of exposure: activity spaces, residential neighborhoods, and self-rated health. *Social Science & Medicine*, 146: 204–213. doi: 10.1016/j.socscimed.2015.10.040

Sun Daosheng, Chai Yanwei, Zhang Yan, 2016. The definition and measurement of community life circle: a case study of Qinghe area in Beijing. *Urban Development Studies*, 23(9): 1–9. (in Chinese)

Sun Daosheng, Chai Yanwei, 2017. Study on the urban community life sphere system and the optimization of public service facilities: a case study of Qinghe area in Beijing. *Urban Development Studies*, 24(9): 7–14, 25. (in Chinese)

Tang L, Xiong C F, Zhang L, 2015. Decision tree method for modeling travel mode switching in a dynamic behavioral process. *Transportation Planning and Technology*, 38(8): 833–850. doi: 10.1080/03081060.2015.1079385

The Central Committee of the Communist Party of China and the State Council of China, 2014. The National New-type Urbanism Plan. http://www.gov.cn/gongbao/content/2014/content_2644805.htm, cited 15 January 2020. (in Chinese)

Therneau T, Atkinson B, Ripley B, 2019. *Rpart: Recursive Partitioning for Classification*. R package version 4.1–15. Available at: https://repo.bppt.go.id/cran/web/packages/rpart/rpart.pdf

Thulin E, Vilhelmson B, Schwanen T, 2020. Absent friends? Smartphones, mediated presence, and the recoupling of online social contact in everyday life. *Annals of the American Association of Geographers*, 110(1): 166–183. doi: 10.1080/24694452.2019.1629868

Tribby C P, Miller H J, Brown B B et al., 2017. Analyzing walking route choice through built environments using random forests and discrete choice techniques. *Environment and Planning B: Urban Analytics and City Science*, 44(6): 1145–1167. doi: 10.1177/0265813516659286

Wang Bo, Zhen Feng, Wei Zongcai et al., 2015. A theoretical framework and methodology for urban activity spatial structure in e-society: empirical evidence for Nanjing City, China.

Chinese Geographical Science, 25(6): 672–683. doi: 10.1007/s11769-015-0751-4

Wang F R, Ross C L, 2018. Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(47): 35–45. doi: 10.1177/0361198118773556

Wang J, Kwan M P, Chai Y W, 2018. An innovative context-based crystal-growth activity space method for environmental exposure assessment: a study using GIS and GPS trajectory data collected in Chicago. *International Journal of Environmental Research and Public Health*, 15(4): 703. doi: 10.3390/ijerph15040703

Wu Qiuqing, 2015. The exploration on the dynamic programming of community in megacities from the living circle perspective. *Shanghai Urban Planning Review*, (4): 13–19. (in Chinese)

Xi G, Zhen F, Cao X et al., 2020b. The interaction between e-shopping and store shopping: empirical evidence from Nanjing, China. *Transportation Letters*, 12(3): 157–165. doi: 10.1080/19427867.2018.1546797

Xi G L, Cao X Y, Zhen F, 2020a. The impacts of same day delivery online shopping on local store shopping in Nanjing, China. *Transportation Research Part A: Policy and Practice*, 136: 35–47. doi: 10.1016/j.tra.2020.03.030

Xiao Jinghao, Zhou Dailin, Hu Jiapei, 2018. Measurement and evaluation method of community life-cycle based on decision tree theory: panyu district of Guangzhou. *Planners*, 34(3): 91–96. (in Chinese)

Xie C, Lu J Y, Parkany E, 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1854(1): 50–61. doi: 10.3141/1854-06

Xu Xiaoyan, Ye Peng, 2010. On the relationship between self-sufficiency and location of urban community facilities. *Urban Problems*, (3): 62–66. (in Chinese)

Yu Yifan, 2019. From traditional residential area planning to neighborhood life circle planning. *City Planning Review*, 43(5): 17–22. (in Chinese)

Zhang M Z, He S J, Zhao P J, 2018. Revisiting inequalities in the commuting burden: institutional constraints and job-housing relationships in Beijing. *Journal of Transport Geography*, 71: 58–71. doi: 10.1016/j.jtrangeo.2018.06.024

Zhen F, Cao Y, Qin X et al., 2017. Delineation of an urban agglomeration boundary based on Sina Weibo microblog ‘check-in’ data: a case study of the Yangtze River Delta. *Cities*, 60: 180–191. doi: 10.1016/j.cities.2016.08.014