Optimizing Bus Stop Spacing Using the Simulated Annealing Algorithm with Spatial Interaction Coverage Model

YUNLEI LIANG, University of Wisconsin, USA
SONG GAO, University of Wisconsin, USA
TIANYU WU, University of Waterloo, Canada
SUJING WANG, University of Waterloo, Canada
YUHAO WU, University of Waterloo, Canada

Public transportation systems, in particular, bus systems, play an essential role in the process of urbanization. Typically more bus stops enable more people to access the bus whereas lower the efficiency of bus system. This study uses a Spatial Interaction Coverage (SIC) model to identify and remove redundant bus stops while maintain the overall success of the whole bus system. The SIC model aims to model the relationship between demand points and bus stops. It takes factors such as the distance and the attractiveness of each bus stop into consideration. The simulated annealing algorithm is then applied with the SIC model to find the optimal combination of bus stops. By applying the SIC model to the iXpress 202 route in Kitchener-Waterloo region, it can effectively identify the number of stops to maintain and remove redundant stops. The bus operation efficiency can be increased by 7.28% after optimization. The SIC model provides a reliable method of modelling the interaction between facilities and demands. The ability of considering the attractiveness of stops and including distance decay into the model can help the transportation agency better plan bus routes. The relationships between bus ridership and the socioeconomic variables (population, income, and age) in the study area are also analyzed.

CCS Concepts: • Information systems → Geographic information systems; • Applied computing → Transportation; • Computing methodologies → Modeling and simulation;

Additional Key Words and Phrases: Optimization, Transportation, GIS, Interaction

ACM Reference Format:

1 INTRODUCTION

1.1 Overall performance of public transportation

Public transportation systems have the carrying capacity to serve large volume of space-time-concentrated travel demand [13]. A successful transportation system can improve the living standard in modern society through reducing
the living expense directly or indirectly[26]. Generally, the overall success of a bus system is determined by the accessibility and the efficiency of the system [7]. The accessibility refers to how easily people can access the bus system [20]. One major factor is the physical distance between potential bus riders and bus stops, as people tend to go to the stop that is nearest to them [20]. With the number of bus stops increasing, the average walking distance to bus stops will decrease, and the system becomes more accessible for riders. The efficiency of a bus system has been defined as how far a bus can reach within a given time window[20]. Given that most people prefer a faster bus, a bus system with high efficiency is more attractive. However, the accessibility and the efficiency cannot be achieved at the same time. To improve the accessibility of the bus system, more stops need to be added, which would lower the speed and raise the operating time of the bus. One way to keep the balance between the accessibility and the efficiency is to optimize bus stop spacing and remove redundant stops [16]. In this way, the bus system will obtain a higher efficiency level while maintain a relatively large coverage of potential riders.

1.2 Approaches to model p-median problems

The process of optimizing bus stop spacing can be modeled as an extension of the p-median problem, which initially aims to “find p facilities to minimize the demand weighted average distance between the demand nodes and the selected facilities” [5]. In this way, the bus stops can be considered as facilities and the goal of optimization is to find p stops that can maximize the overall success of the bus system [7, 24]. There are a few major concepts discussed in previous literatures: the willingness-to-walk distance [9, 17], and distance decay[12]. The willingness-to-walk distance has been defined as the the maximum distance that people are willing to travel from their origins to bus stops[4, 21]. A 400-meter distance [22] or a 5-minute walk is known as a typically used value[20]. The distance decay describes the decreasing likelihood that people would take the bus with the increasing distance[9, 20].

Studies have combined multiple concepts in their models to construct the objective functions of the bus stop spacing optimization problem. A Spatial Interaction Coverage (SIC) model was constructed by modifying the Huff model [7]. The Huff Model addresses that the probability of a trip is positively related to the relative attractiveness of the destination, and inversely related to the relative distance to the destination compared with its competitors [15]. Both the cost of access and the attractiveness are included in the SIC model. The cost of access is determined by the walking distance, and the attractiveness is defined by the surrounding environment (e.g. the existence of shelters, distance to shopping centres) and the number of bus routes at one stop[3]. Given a number of stops p, the optimal solution is considered to be the one with the largest spatial interaction coverage.

In this work, we presents a modified SIC model combined with the simulated annealing (SA) algorithm to optimize the bus stop spacing. The following of the paper is organized as follows. Section 2 introduces our study area, data preparation and processing workflow, as well as our spatial optimization approach to solving this problem. Section 3 presents the bus stop selection and spacing optimization results. Section 4 then discusses the broad implications of this study. Finally, section 5 concludes this work.

2 METHODS

2.1 Study area and Data

The study area is the Kitchener and Waterloo (KW) region of the Ontario Province in Canada. According to the 2016 census data of Canada, the population of KW region was 338,208[1]. The total area of this region is 202.96 km² [8].
Optimizing Bus Stop Spacing

The main public transportation authority in the KW region, Grand River Transit (GRT), operates 32 local routes and 5 express routes that connect approximately 2,600 stops in this region (GRT, 2017). The average annual ridership of GRT is around 19.7 million and the on-time rate of buses is about 80% [11]. There are three universities residing in this area: Conestoga College, the University of Waterloo, and the Wilfrid Laurier University. With three universities in this area, taking GRT buses has become a desirable way to transfer for students and locals in this region.

The iXpress 202 bus route (Figure 1) is used as a case study. The route runs from the Boardwalk Terminal station in the southwest to the Conestoga Mall terminal station in the north. This route contains 23 stops, with a total length of approximately 18 km. The average speed of the bus is around 27 km/h and the total operating time is around 40 minutes. Data of bus routes and stops in the KW region are provided by the Region of Waterloo Open Data portal. The ridership data of iXpress 202 in fall 2017 is obtained from GRT. Apart from the ridership data, the census population at the dissemination block (DB) level and the household income of private households before tax and median age at dissemination area (DA) level are obtained from the Statistics Canada. In addition, the data of academic and commercial facilities are provided by the Region of Waterloo Open Data portal.

The overall process of the study can be divided into four steps: 1) preparing and processing data; 2) calibrating parameters for SIC model and SA algorithm; 3) implementing the SA algorithm to solve the optimization problem based on SIC; 4) presenting the optimization results and analyzing the reasons why certain stops are removed.

2.2 Data Processing

The main input of the SIC model is the distance and the attractiveness of each bus stop. According to the professional suggestion from GRT, a bus stop should be accessible within around 450 meters from the residing location. The centroid of each dissemination block are used as riders’ origins, which may increase the distance from some points to the stop.
Therefore, this study chose 500 meters as the standard willingness-to-walk distance. And the population, average income and median age data were collected to estimate the ridership. Before the construction of the SIC model, the attractiveness of each stop was calculated based on the connectivity, surrounding and the infrastructure of bus stops. Specifically, the number of bus routes at each stop is used to represent its connectivity, and a stop with more stopover routes will have a higher attractiveness value. The existence of malls and universities nearby (i.e. 500m) bus stops will be considered as the surrounding variables. If a stop is located within a 500-meter buffer of a mall or an university, it has a high attractiveness. The infrastructure represents shelter, electronic notice boards, or seats located at the bus stop. In this study, only shelters are considered. The analytic hierarchy process (AHP) was used to calculate the attractiveness score for each stop. It is a multi-criteria decision-making approach that can assign weights to each criterion by paired comparisons[23]. Weights of connectivity, surrounding of each bus stop and shelter infrastructure are shown in Table 1.

<table>
<thead>
<tr>
<th>Connectivity</th>
<th>Shopping Mall</th>
<th>University</th>
<th>Shelters</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.36</td>
<td>0.28</td>
<td>0.27</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 1. The weight of each factor that influences bus stop attractiveness.

2.3 Construction of the objective function

The equations of SIC and definitions of variables are shown in the following. The objective of the optimization problem can be constructed based on SIC, and the corresponding objective function is to maximize the total spatial coverage $Z$ with constraints (equation (1)-(2)).

Maximize:

$$Z = \sum_{i \in I} \sum_{j \in J} S_{ij}$$

$$S_{ij} = \left[ \frac{w_j^\alpha d_{ij}^{-\beta}}{\sum_{k \in N_i} w_k^\alpha d_{ik}^{-\beta}} \right] \alpha_i X_j$$

where:

- $i$ = index of demand nodes
- $j, k$ = index of bus stops
- $I$ = collection of demand nodes
- $J$ = collection of all bus stops
- $d_{ij}$ = shortest accessing distance between demand node $i$ and candidate bus stop $j$
- $w_j$ = attractiveness of bus stop $j$
- $\alpha_i$ = demand at location $i$
- $\alpha$ = exponent that controls $w_j$
- $\beta$ = exponent that controls $d_{ij}$
- $N_i$ = a set of candidate bus stops within the willingness-to-walk distance
- $X_j = \begin{cases} 1 & \text{if existing candidate bus stop $j$ is selected} \\ 0 & \text{otherwise} \end{cases}$
- $S_{ij}$ = spatial interaction between demand node $i$ and candidate bus stop $j$
Optimizing Bus Stop Spacing

The SIC model assumes that the number of riders in a demand node can be allocated to different bus stops that are within the willingness-to-walk distance[7]. For each stop, the proportion of allocated demand is positively affected by the relative attractiveness of the given stop compared to other candidate stops. The shortest accessing distance to the given stop compared to others has a negative influence (equation (2)). Two parameters $\alpha$ and $\beta$ are used to control the weight of attractiveness and accessing distance in the SIC model.

To find the parameters that can best represent the situation of this study area, the performance of SIC with different parameters is assessed by comparing the SIC estimated ridership with the actual ridership. The ridership of a bus stop ($i$) is considered as a dependent variable $G_i$ predicted by three independent attributes (i.e. population ($P$), income ($I$), and age ($A$)). The population, income and age data of each stop is aggregated from the demand points with the proportion calculated by the SIC model.

$$G_i = a \cdot P_i + b \cdot I_i + c \cdot A_i$$

(3)

Iterative tests with different sets of SIC parameters are implemented. The initial values of $\alpha$ and $\beta$ are selected randomly. With the current set of parameters, a SIC equation can then be uniquely determined. By looping through the OD Matrix, the three DB attributes of each origin in the matrix are aggregated to each stop based on the determined SIC model. The linear regression is then applied to determine the coefficients of each factor, which constructs the ridership estimation function. The adjusted $R^2$ indicates the fitness between the regression function and actual ridership data and is used to determine the performance of the SIC model.

After a number of trials, the sets of parameters are compared based on the adjust $R^2$ and the one with the highest $R^2$ is chosen as the appropriate parameter pair for the study area, and the corresponding ridership estimation function is generated.

2.4 The SA algorithm

Bus stop spacing optimization is a kind of p-median problem with NP hardness, and this means that the computing time of solving a specific p-median problem can increase extremely when the value of $p$ increases [10]. Several heuristic algorithms such as the tabu searching algorithm, simulated annealing (SA) algorithm and genetic algorithm have been introduced to facilitate solving the p-median problems (Fathali, 2006). The SA algorithm originates from the cooling process and has been widely adopted to solve optimization problems because of its concise process, high time efficiency and high accuracy [6, 18, 19]. The SA algorithm is used in our study to solve the optimization problem constructed by the modified SIC model. The efficiency of SA is dependent on both the initial solution and choice of parameters[18]. The generation of the initial solution is based on random selection when the algorithm is firstly introduced[2]. Currently, there is no standard for the choice of SA parameters, and an appropriate initial solution varies among different optimization problems. Therefore, it is recommended to make guesses or use a trial-and-error method to determine the initial solution and other parameters of SA[2]. This study randomly chose $p$ bus stops as the initial solution. The SA parameters including initial temperature, cooling rate and minimum temperature are determined arbitrarily at first, and by comparing the time cost and the solution variance of different parameter pairs, those parameters that make SA most efficient and suitable for the optimization problem are adopted.

The optimization process consists of two steps. First, the maximized spatial interaction coverage of different number of bus stops will be calculated, and based on the spatial interaction coverage and its decreasing rate, a number $p$ as the minimum number of stops to maintain the service success of the bus route will be recognized. Second, SA will be used to...
to find the specific $p$ stops that should be retained to maximize the spatial interaction, and to remove any unnecessary stops.

3 RESULTS

3.1 Linear Regression

3.1.1 Choosing the parameters for the SIC model. The linear regression was completed using the ridership data in fall 2017 provided by the local transit authority. The linear regression fits the estimated population, income and age data with actual ridership data. The results are used to determine appropriate parameters of SIC (i.e. $\alpha$ and $\beta$) as well as the coefficients of the ridership estimation function. The performance of difference $\alpha$ and $\beta$ is determined by the adjust $R^2$ and the $p$ value. The adjust $R^2$ indicates how the data fits with the model. With a higher $R^2$, the data fits better with the model. The $p$ value can represent how such goodness of fit is statistically significant.

In our parameter calibration process, the $\alpha$ and $\beta$ values change from 0.1 to 4 with a step of 0.1. Among all the chosen parameters, the pair of $\alpha = 1$ and $\beta = 2$ shows a highest adjust $R^2$ of 0.8083; therefore, it is chosen as the final parameter pair of the SIC model.

3.1.2 The choice of the aggregation method. When allocating the demand points to different stops using the SIC model, this study allocates the population, income and age data by proportion to each stop. Then, for each stop, its total population, income and age data is obtained by aggregating the population, income and age data allocated from each demand point. Then there is a problem about how to aggregate the population, income and age data. To be clear, whether the total population or other metrics calculated from the demand points should be used to represent the population of a bus stop. For each attribute (population, age and income), there are two choices: the average or the total.

Therefore, the total eight combination were tested. For every aggregate method, the allocation result is compared with the actual ridership data and the adjusted $R^2$ is used to decide which method is the best.

<table>
<thead>
<tr>
<th></th>
<th>TAA</th>
<th>TTT</th>
<th>AAA</th>
<th>TTA</th>
<th>ATA</th>
<th>ATT</th>
<th>AAT</th>
<th>TAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.8083</td>
<td>0.07179</td>
<td>0.7612</td>
<td>0.792</td>
<td>0.5616</td>
<td>0.09713</td>
<td>0.5144</td>
<td>0.4753</td>
</tr>
</tbody>
</table>

Table 2. The adjusted $R^2$ value of eight aggregate methods (T means Total and A means Average. The default sequence is Population, Income, Age).

From Table 2, the combination TAA has the highest adjusted $R^2$, which means the allocation result fits the actual ridership data the best. Therefore, the combination of total population, average income and average age was used to aggregate the demand points to each stop.

3.2 Determining the number of stops to retain using the SIC model

The maximized spatial interaction coverages of different numbers of bus stops were calculated using SA, and ten times of repetitive calculations were executed to examine the variance of different solutions. To demonstrate the decreasing trend of spatial interaction coverage, all the maximized spatial interaction coverage with different number of stops were normalized as the percentage of the initial value. The plot of percentage against the numbers of stops are shown in Figure 2. The curve indicates that the spatial interaction coverage increases gradually when the number of stops increases. But the increasing rate is becoming slower with more stops. The spatial interaction coverage value rises slightly when the stop number is over 20. Based on this figure, the minimum number of stops to be retained is determined as 19, given that 19 stops can yield relative high spatial interaction coverage (0.95 of initial spatial interaction coverage).

Manuscript submitted to ACM
3.3 Overall distribution of the removed stops

Figure 3 shows the 500-meter and 1000-meter service areas of the retained stops (blue) and the removed stops (yellow) after optimization. The four removed stops are located at the northeast region of the study area. The removed stops have relatively low rank of weekday ridership ranging from 11 to 17 (Table 3). For the attractiveness, all stops have an attractiveness value of 3, which is the second lowest attractiveness rank. With such low ridership and low attractiveness, it is reasonable that those four stops were removed.

![Service area of retained and removed stops using the SIC model.](Image)

<table>
<thead>
<tr>
<th>Stop ID</th>
<th>Weekday Ridership</th>
<th>Weekday Ridership Rank</th>
<th>Attractiveness</th>
<th>Attractiveness Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>3671</td>
<td>74.8</td>
<td>15 out of 23</td>
<td>3</td>
<td>6 out of 7</td>
</tr>
<tr>
<td>3766</td>
<td>56.8</td>
<td>17 out of 23</td>
<td>3</td>
<td>6 out of 7</td>
</tr>
<tr>
<td>4049</td>
<td>144.5</td>
<td>11 out of 23</td>
<td>3</td>
<td>6 out of 7</td>
</tr>
<tr>
<td>4044</td>
<td>68.7</td>
<td>16 out of 23</td>
<td>3</td>
<td>6 out of 7</td>
</tr>
</tbody>
</table>

Table 3. The ridership, attractiveness and their ranking of removed stops of the SIC model.

3.4 Hot-spot analysis

To better illustrate why certain stops are removed, the hot-spot maps based on different factors are created. As an example, figure 4 shows the hot-spot maps of the average income. Three out of four removed stops (cyan) fall in red regions, where exists significant spatial clusters of high income. With high income, the possibility of owning private cars will increase and this may lead to the decreasing demand for public transportation[14]. The correlation analysis reveals that the ridership negatively correlates with the average income of the bus service areas with a Pearson’s coefficient of -0.76.
3.5 Efficiency Improvement

Using the introduced spatial optimization model, the operating time can be reduced by approximately 1.67 minutes per ride after removing 4 bus stops (25s per stop). There are 47 bus rides per weekday, thus, 1.31 hours can be saved per day. The efficiency can be increased by 7.28% after optimization. There are two periods of rush hour per day, and the saved time can be used to add one more bus ride during each period. The time interval in the morning is 13.17 minutes and in the evening is 14 minutes before optimization. After adding one bus ride for each period, the time interval can be reduced by 1.88 minutes and 53 seconds for the morning and evening period respectively (Table 4).

<table>
<thead>
<tr>
<th>Rush Hour</th>
<th>Morning</th>
<th>Evening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time interval before optimization</td>
<td>0:13:10</td>
<td>0:14:00</td>
</tr>
<tr>
<td>Time interval after optimization</td>
<td>0:11:17</td>
<td>0:13:07</td>
</tr>
<tr>
<td>Reduced time interval</td>
<td>0:01:53</td>
<td>0:00:53</td>
</tr>
</tbody>
</table>

Table 4. The reduced time interval of two periods after optimization.

4 DISCUSSIONS

Figure 5 shows the histogram of income of all demand points covered by the service area of all existing bus stops. The demand points that fall in the service areas of four removed stops are selected and they fall in the highlighted bars. All of the selected demand points have annual average income larger than 25,000 dollars. According to Canadian Socio-Economic Information Management System[25], the cut-off of low income before tax for the community that has a population over 500,000 (similar to the size of the KW region) is about 19,307 dollars. Based on this information, people who live in the coverage area of removed stops have relatively good living conditions. Many previous studies stated that a high income can lead to decreasing demand for public transportation.

The multiple linear regression model is further constructed using three independent variables including total population, average income, and average age. The dependent variable is the ridership of each bus stop. Table 5 shows the statistics of the regression result. Based on the Standardized Regression Coefficients, all three independent variables are all
negatively related with the ridership. Changing in age is significantly related to changes in ridership while changes in population and income have less significant effects on the ridership. For the whole model, the p-value of the F-statistic is less than 7.926e-05, which indicates that the model is highly significant. Note that the statistical results are only drawn from our case study data though. The findings of this study may compensate for the limitations of existing studies [7], where the income is not the dominated factor.

This study provides a reliable and efficient approach to identify the number of redundant bus stops in the bus route considering the bus attractiveness and the walking distance of riders. This study also integrates linear regression with the actual ridership data to modify the weights of allocation variables as well as construct the ridership estimation function. This ensures a better ridership estimation that can simulate the actual situation. If researchers can identify the minimal number of stops to be retained, it is possible to predict the ranges of future ridership changes based on the calculated function. Hence, this approach can be used to improve the design of existing bus routes, specifically for the design of express routes, whose overall success is highly emphasized.

5 CONCLUSIONS

This study presents a modified SIC model combined with an optimization SA algorithm to identify the redundant stops of a bus route. With a goal to maximize the spatial interaction coverage, the study firstly ran the model to obtain the maximum SIC value with different number of bus stops. After the number of stops that can maintain a relative high coverage while remove several unnecessary stops was determined, the specific bus stops were selected by rerunning the model with the chosen stop number. The bus operation efficiency can be increased by 7.28% after optimization in our case study. The main contribution of this study to the research of bus stop optimization lies in the modification of SIC parameters, namely, the construction of ridership estimation function based on linear regression. By adjusting the parameters using the ridership data rather than randomly selecting the parameters, the model can better reflect the actual situation and provide more reliable strategies to remove redundant stops. In future work, we would like to apply our approach in multiple bus routes and develop scalable spatial optimization framework for optimizing the city-scale bus operation stops. In addition, the satisfaction of bus passengers after the efficiency optimization requires further investigation.

Table 5. The statistics of the linear regression model

|          | Estimate | Std. Error | Pr(>|t|) | Standardized Regression Coefficient |
|----------|----------|------------|----------|-----------------------------------|
| (Intercept) | 903.741297 | 2.15e-06 *** | 0.34864 | -0.1969802 |
| Pop       | -0.115808 | 0.07867 .   | -0.3344371 |
| Income    | -0.001895 | 0.00106 **  | -0.6999719 |
| Age       | -15.228742| 0.00106 **  | -0.6999719 |

Fig. 5. The histogram of Income of all Demand Points.
REFERENCES