Research Article
Urban Impedance Computing Based on Check-In Records

Zhiyong Yu,[1] Yuzhong Chen,[1] Songpan Zheng,[1] Yao Shen,[2] Zhiwen Yu,[3] and Jordan Pascual[4]

[1]*College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China*
[2]*Space Syntax Lab, Bartlett School of Architecture, University College London, London NW1 2BX, UK*
[3]*College of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China*
[4]*Department of Computer Science, University of Oviedo, 33024 Gijón, Spain*

Correspondence should be addressed to Zhiyong Yu; yuzhiy@gmail.com

Urban impedance is an important consideration in assessments of transportation and land-use systems. This work leverages check-in records obtained from mobile social networks to build a fine-grained but inexpensive urban impedance model. Check-in records and road networks are collected and used to calculate and adjust the various parameters of the model, including path length, number and angle of turns, number and direction of junctions, and population density. Check-in records can filter functional locations and supply the time factor, thereby providing excellent advantages over traditional models that do not employ this data type. The proposed model is more accurate than traditional impedance models, as verified by experiments using Sina Weibo data in Tianjin City.

1. Introduction

The rapid progress of urbanization has modernized many people's lives but also engendered several important issues, such as traffic congestion, high housing prices, and low neighborhood cohesion. Urban accessibility [1] is a fundamental measure of the benefits of urban life and answers how many destinations can be accessed within a given time. It is widely used to evaluate transportation and land-use systems. Increasing accessibility is essential for urban planning. Impedance, which is defined as the cost between the origin and destination locations, is an important indicator of urban accessibility.

Traditional impedance models require a large amount of supporting data, including the geographical distribution of public facilities, traffic volume of different time periods, and population in the range of study [2]. These data are difficult to collect from many cities of developing countries. As the use of mobile social networks has recently grown immensely, a large amount of user mobility data has been generated. Mobile social networks allow a user to "check in" at POIs (points of interest), which corresponds to an online record describing his/her current physical location, and share this record with his/her friends. This new data source can be leveraged to build a fine-grained but inexpensive urban impedance model. The present paper explores this idea and presents the following main contributions:

(1) Selection and preparation of appropriate data sources for the impedance model: we extract city blocks from road networks, collect POIs and check-ins from mobile social networks, and match them correspondingly.

(2) Proposal and implementation of the impedance model: we use check-in records to filter POIs and supply the time factor and then calculate distances, turns, junctions, and population parameters after path planning.

(3) Visualization of impedance and its parameters in a real city.

2. Related Work

In recent years, researchers have studied a number of urban accessibility models. Handy and Niemeier [3] proposed a utility model based on the discrete-choice model. By assessing a value for each candidate destination in the region, individuals will be more likely to visit destinations with larger utilities. Wachs and Kumagai [4] proposed a cumulative opportunity model that considers the number of destinations in a certain range; unfortunately, this model ignores differences in opportunity points. Hansen [5] proposed a gravity model that considers the distance between centers and opportunity points and could also be used to analyze the market potential [6]. Accessibility is heavily influenced by the selection of the impedance function. Common impedance functions include exponential functions, Gaussian functions, and negative power functions. The choice of impedance function mainly depends on the applications of the model and the characteristics of the data used. Since their functions and parameters are simplified, these models tend to ignore the complexity of city roads and, therefore, show limited accuracy.

In research on impedance functions, the United States BPR (Bureau of Public Roads) developed [7] a speed-flow model through regression analysis. To overcome the shortcomings of BPR model resulting in large deviations when the traffic volume reaches its peak, Spiess [8] and Wang et al. [9] improved the impedance function. According to queuing theory, Davidson [10] proposed a progressive impedance function that amends the parameters describing the relationship between travel time and distance. Wang et al. [11] improved the impedance model based on traffic flow by considering travel costs, time, hub locations, and other factors. Despite their benefits, however, as mentioned earlier, these traditional impedance models require a significant amount of supporting data.

Another research field related to urban accessibility is urban computing or smart city. Urban computing [12] aims to tackle urbanization issues through a process involving acquisition, integration, and analysis of data generated from various sources in urban spaces, such as sensors, devices, vehicles, buildings, and humans, to reflect traffic flows, human mobility, geographical information, and so forth. For example, Zheng et al. [13] described the underlying problems in Beijing's transportation network using the hypothesis that the connection between two regions cannot effectively support the traffic traveling between them, thereby resulting in a large volume, low speed, and high detour ratio. Fu et al. [14] predicted the rankings of residential real estate in a city at a future time according to their potential values inferred from a variety of data sources, such as human mobility data and urban geography, currently observed around the real estate. Chen et al. [15] leveraged a combination of location-based social networks and taxi GPS digital footprints to achieve personalized, interactive, and traffic-aware trip planning. Yu et al. [16] recommended personalized travel package with multiple points of interest based on crowd sourced user footprints. These works are mostly application-oriented and can be embedded into and improved by our model.

3. Urban Impedance Model

3.1. Preliminary

Definition 1 (block center). Road networks divide the city area into polygonal blocks [17, 18]. Let $V = \{v_1, v_2, \ldots, v_{|V|}\}$ indicate the centers of these blocks and let $|V|$ indicate the number of blocks available. People and facilities in a block usually share the same transport environment; thus, block centers may be considered representative locations when analyzing the impedance of a block (we can compute the impedance for all locations, but doing so is time-consuming and unnecessary). The position of the center of a block can be easily obtained given its vertices and edges.

Definition 2 (function density). This term considers how many functional locations (i.e., POIs) can be reached from an object location (e.g., the center of a block) given an accessibility radius. The positions of POIs are indicated by longitude and latitude. The function density is displayed in formula (1). Consider

$$\text{Den}_i = \sum_{j=1}^{k} \partial_{ij}, \quad \partial_{ij} = \begin{cases} 1, & d\left(f_j, v_i\right) \leq r \\ 0, & d\left(f_j, v_i\right) > r, \end{cases} \quad (1)$$

where $d(f_j, v_i)$ denotes the road network distance between a POI f_j and the center of a block v_i, r is the accessibility radius defined in terms of road network distance (because people travel along road networks and care about transport distances rather than absolute distances), and k is the total number of POIs.

POIs, such as shops, restaurants, and tourist attractions, have business hours, for example, opening from 8:00 a.m. to 8:00 p.m. Thus, a time factor must be considered when calculating their impedance. Out of their business hours, these POIs are regarded as absent. We set a time slot lasting for two hours and classify days into weekdays or weekends. As a result, a total of 24 time intervals should be considered. For each block center, the function density is adjusted according to formula (2). Consider

$$\text{Den}_{i,t} = \sum_{j=1}^{k} \partial_{ij} \beta_{jt},$$

$$\partial_{ij} = \begin{cases} 1, & d\left(f_j, v_i\right) \leq r \\ 0, & d\left(f_j, v_i\right) > r \end{cases}, \quad \beta_{jt} = \begin{cases} 1, & t \in \text{time}\left(f_j\right) \\ 0, & t \notin \text{time}\left(f_j\right), \end{cases} \quad (2)$$

where $\text{time}(f_j)$ denotes the business hours of a POI f_j, t is a specific time slot, and the other variables are identical to those in formula (1). POIs with ≥ 10 check-ins are taken into account in this work.

Definition 3 (path). A path is a part of road network that starts from an origin location and ends at a destination location. A path is described by five elements $\langle O, D, \mathcal{T}, \mathcal{J}, \mathcal{L} \rangle$: origin ($O$), destination ($D$), turns ($\mathcal{T}$), junctions ($\mathcal{J}$), and line segments among them ($\mathcal{L}$). Consider $\mathcal{T} = \{\theta_1, \theta_2, \ldots, \theta_{|\mathcal{T}|}\}$,

where $|\mathcal{T}|$ is the number of turns and θ_i is the angle of each turn. Consider $\mathcal{J} = \{c_1, c_2, \ldots, c_{|\mathcal{J}|}\}$, where $|\mathcal{J}|$ is the number of junctions and c_i is the direction to take upon arriving at a junction, that is, turning right, turning left, or going straight. Consider $\mathcal{L} = \{d_1, d_2, \ldots, d_{|\mathcal{T} \cup \mathcal{J}|+1}\}$, where $|\mathcal{T} \cup \mathcal{J}|$ is the total number of turns and junctions and d_i is the length of each line segment.

We adopt the Baidu map traffic routing API [19] to obtain paths between block centers and POIs.

3.2. Modeling. The overall impedance situation of a city can be understood by the impedances of its block centers, which are denoted by $\{m_1, m_2, \ldots, m_{|V|}\}$. The larger the value of m_i, the larger the impedance of this block center.

Drawing lessons from existing models, we take both path parameters and population parameters into consideration. We also use check-in records to adjust these parameters. Check-in records can help our modeling procedure in two ways. First, check-in records can filter functional locations. Only locations that own many/frequent check-ins are meaningful for impedance computing. Second, check-in records can supply the time factor. A location presents different impedances at different times. Time not only influences the function density (as shown in formula (2)) but also impacts the real-time population (which will be described in Section 3.3.4).

Three parameters are considered for paths, that is, path length (L), number and angle of turns (T), and number and direction of junctions (J), and one parameter is considered for population, that is, the population density (P). The impedance of the block center v_i at the time interval t is

$$m_{i,t} = \alpha \overline{L_{i,t}} + \beta \overline{T_{i,t}} + \gamma \overline{J_{i,t}} + \delta \overline{P_{i,t}}, \tag{3}$$

where α, β, γ, and δ are the weights of different parameters.

3.3. Parameter Calculation

3.3.1. Path Length. Path length (i.e., road network distance), the most important component of the impedance model, refers to the sum of lengths of line segments in a path. The average of all path lengths from a block center to all accessible POIs is taken as an initial parameter of our model:

$$L_{i,j} = \sum_{k=1}^{|\mathcal{T} \cup \mathcal{J}|+1} d_k,$$

$$\text{where } d_k \in \mathcal{L} \text{ of path from } v_i \text{ to } f_j, \tag{4}$$

$$\overline{L_{i,t}} = \frac{\sum_{j=1}^{\text{Den}_{i,t}} L_{i,j}}{\text{Den}_{i,t}}.$$

We use the CartoDB map tool [20] for visualization. Figure 1 shows the average path lengths of Tianjin's block centers (at a particular time interval). Blocks with darker colors represent blocks with longer average path lengths. A detailed description of the data will be provided in Section 4.
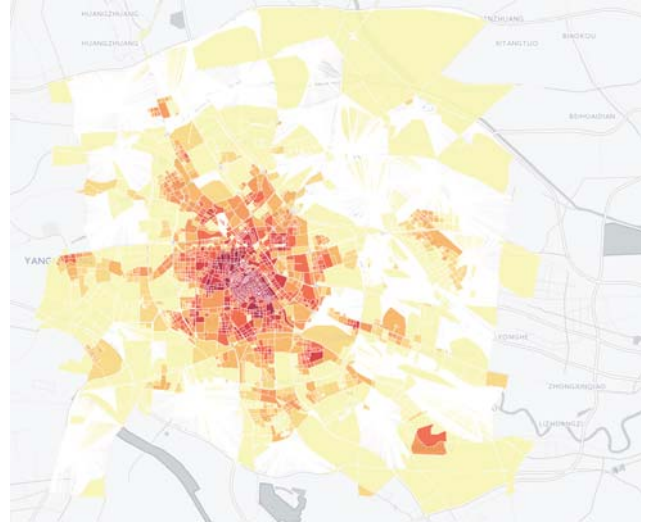

FIGURE 1: Average path lengths of Tianjin's block centers.

3.3.2. Number and Angle of Turns. The path from the block center to a POI may contain turns. Obviously, more and sharper turns will slow down traffic. Thus, the number and angle of turns in a path must be taken into account in the urban impedance model.

Baidu map traffic routing indicates a very large number of slight turns. Because modeling all of these turns is time-consuming and unnecessary, angles less than 10 degrees are neglected. Traditional impedance models tend to ignore the effect of the segment between two turns. In practice, however, if only a short distance must be traveled from one turn to another, the whole passing time will be greater. The average of path *turns* (short for the number and angle of turns, the same below) from a block center to all accessible POIs is taken as the second parameter of our model. Consider

$$T_{i,j} = \sum_{k=1}^{|\mathcal{T}|} \ln\left(\frac{\tan(\theta_k/4)}{d_k} + 1 \right),$$

$$d_k \in \mathcal{L}, \ \theta_k \in \mathcal{T} \text{ of path from } v_i \text{ to } f_j, \ \theta_k \geq 10°, \tag{5}$$

$$\overline{T_{i,t}} = \frac{\sum_{j=1}^{\text{Den}_{i,t}} T_{i,j}}{\text{Den}_{i,t}},$$

where d_k is the length of the line segment between the k-1th turn and the kth turn. The function tan() can show angle differences properly, and the function ln() can restrict the value range. Figure 2 shows the average path turns of Tianjin's block centers.

3.3.3. Number and Direction of Junctions. Junctions on a path increase the passing time because people must wait for traffic lights to change and pedestrians from other directions to cross roads. The more the junctions on a path, the larger its impedance. Different directions at a junction present different waiting and passing times. Turning left, for example, requires more cost than going straight and then turning right.
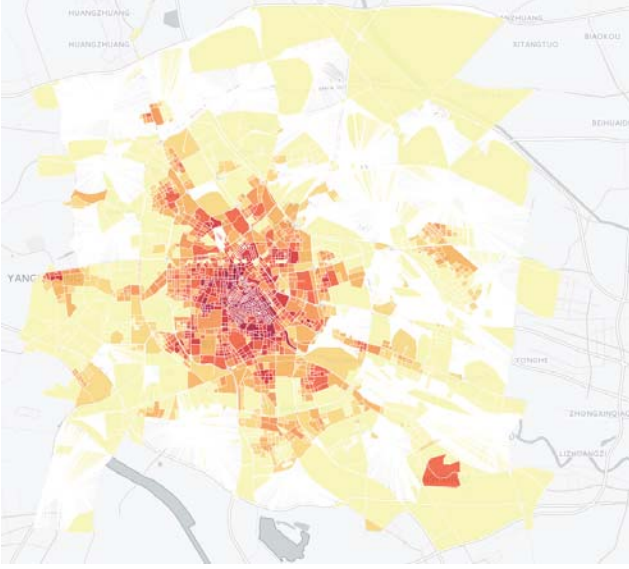
FIGURE 2: Average path turns of Tianjin's block centers.

FIGURE 3: Average path junctions of Tianjin's block centers.

No impedance function is widely used to consider junctions in existing works. In our model, we construct one using logical rules. The average of path *junctions* from block center to all accessible POIs is taken as the third parameter of our model. Consider

$$J_{i,j} = \sum_{k=1}^{|\mathcal{J}|} \frac{c_k}{d_k},$$

$$c_k = \begin{cases} 1, & \text{turn right} \\ 2, & \text{go staight} \\ 3, & \text{turn left} \end{cases}, \; c_k \in \mathcal{J} \text{ of path from } v_i \text{ to } f_j, \quad (6)$$

$$\overline{J_{i,t}} = \frac{\sum_{j=1}^{\text{Den}_{i,t}} J_{i,j}}{\text{Den}_{i,t}},$$

where d_k is the length of the line segment between the k-1th junction and the kth junction. Figure 3 shows the average of path junctions of Tianjin's block centers.

3.3.4. Population Density. The population density directly affects the traffic situation in a block; thus, its impact on impedance cannot be ignored. Different times of day show different population densities. Although some traditional impedance models also consider population, real-time population distribution data are difficult to obtain and their precisions are not able to meet the demands of accurate calculations. Users' check-in records can properly represent the population distribution and can be easily obtained from mobile social networks. As different area sizes present different abilities to relieve traffic congestion, the population density of an approximately circular area determined by
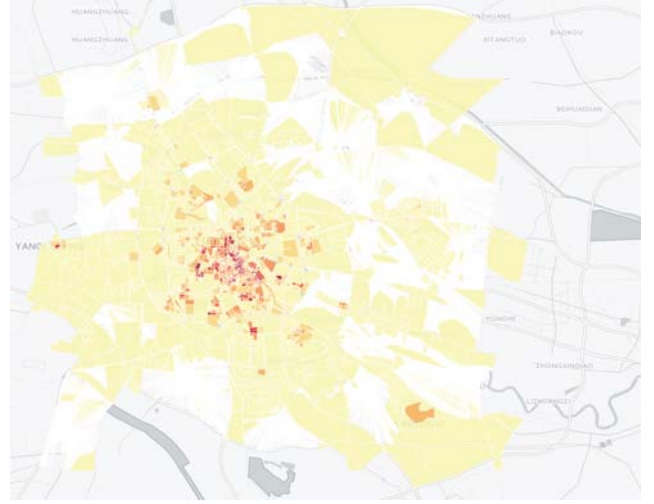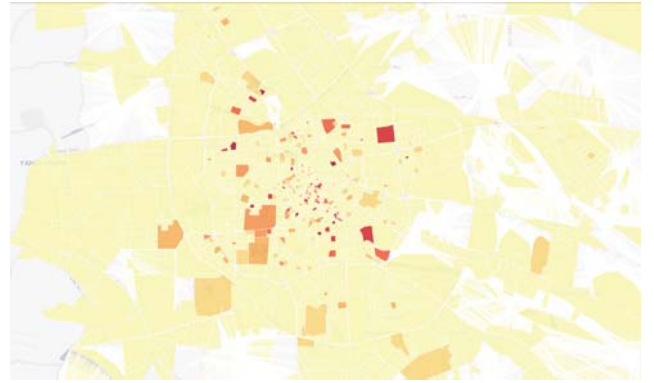

FIGURE 4: Population density of Tianjin's block centers (8:00 a.m.–10:00 a.m., weekend).

the block center and the accessibility radius is taken as the fourth parameter of our model:

$$\overline{P_{i,t}} = \frac{\sum_{j=1}^{\text{Den}_{i,t}} \text{check}_{j,t}}{s_{i,r}}, \quad (7)$$

where $\text{check}_{j,t}$ is the number of check-ins at POI f_j and $s_{i,r}$ is the size of an approximately circular area with center v_i and accessibility radius r (simply calculated as πr^2 in this work). Figure 4 shows the population density of Tianjin's block centers (from 8:00 a.m. to 10:00 a.m. on a weekend).

4. Experiments

The experimental data include road networks from the national data center and POIs and check-in records from Sina Weibo [21]. We use review records of POIs from Dianping [22] for cross-checking, as well as two traditional impedance models (the potential and utility models) for comparison. Real-time road conditions from the Baidu map are considered the ground truth.
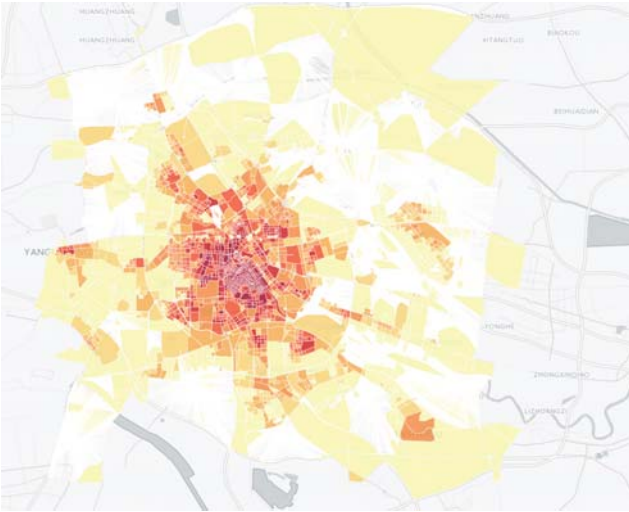
FIGURE 5: Impedance of Tianjin's block centers (8:00 a.m.–10:00 a.m., weekend based on Weibo).
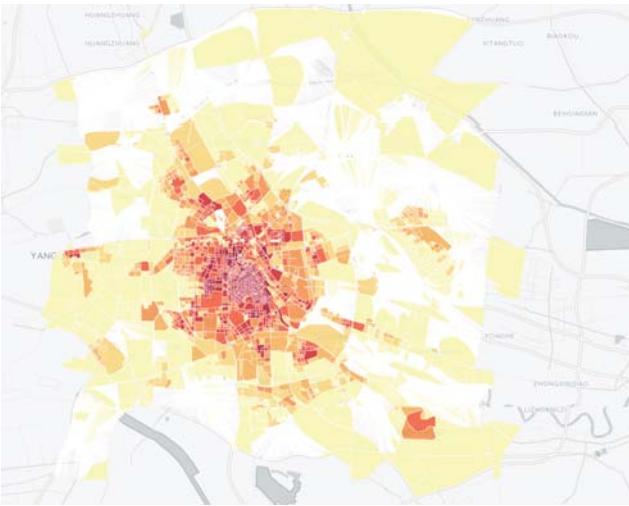

FIGURE 6: Impedance of Tianjin's block centers (8:00 a.m.–10:00 a.m., weekend based on Dianping).

This paper describes the case of Tianjin (one of four municipalities in China) as an example. Road networks divide the city area into 2,754 blocks. A total of 90,731 POIs and 533,006 check-in records were collected in 2014. The accessibility radius was set to 1,000 meters in terms of road network distance, and 357,681 paths should be calculated.

4.1. Cross-Checking with Different Data. First, we present the results according to Weibo data. Figure 5 shows the impedance of Tianjin's block centers (8:00 a.m.–10:00 a.m., weekend).

Using the same impedance model (see formula (3)), we compute Tianjin's impedance according to Dianping data and display the results in Figure 6. Dianping data include 45,767 POIs and 311,794 review records (regarded as check-in records). For the same time interval, the results of
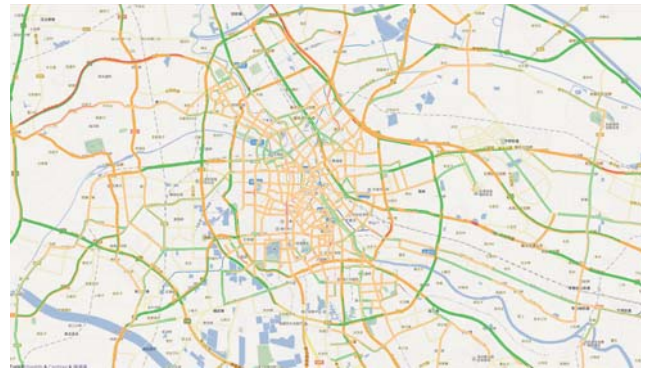

FIGURE 7: Real time road conditions in Tianjin at 9:00 a.m. on a Saturday.

the Dianping data are basically identical to those of the Weibo data. This finding shows that our model works robustly and steadily with different data sources.

4.2. Comparison with Other Models. The utility and potential models are two common impedance models.

The utility model is based on the discrete-choice model. The basic idea of this model is that different facilities present different utilities to people. For example, a supermarket is more likely to be visited than a car shop. For each destination within the accessibility radius, a utility value is assigned to the origin. The greater the utility is, the greater expectation this destination will be visited. The impedance function of this model is described as a logarithmic sum: $m_i = \ln^{-1} \sum_{j=1}^{\text{Den}_i} \exp(u_{i,j})$, where $u_{i,j}$ is the utility value of v_i assigned by f_j. We use the number of ODs (a pair of check-in records left by the same user in a day) as the utility value.

The potential model is based on Newton's gravity model. Here, two factors are considered: the attractiveness of the destination, usually indicated by the population or facility density at the destination, and the distance decay. A typical expression of the potential model is $m_i = \sum_{j=1}^{\text{Den}_i} (\exp(2L_{i,j})/\text{check}_j)$, where f_j's check-in number check_j indicates the attractiveness of f_j.

Real-time road conditions are a good reflection of impedance. We take real-time road conditions from the Baidu map as a reference to evaluate all three impedance models. Figure 7 shows the real-time road conditions in Tianjin at 9:00 a.m. on a Saturday. Here, green lines indicate clear roads, yellow lines indicate slow roads, and red lines indicate roads with traffic congestion. To achieve a more intuitive contrast among the results of the impedance models, we project the road conditions to the corresponding blocks. Figure 8 shows the projected result of road conditions in Tianjin (average from 8:00 a.m. to 10:00 a.m. on a weekend). A comparison of Figures 5 and 8 indicates nearly coincident results.

We now present a quantitative comparison to demonstrate which model yields results most similar to those of the reference (i.e., projected results of real road conditions). The Minkowski distance, a proper similarity criterion, is written

Table 1: Similarity between each model and road condition (weekends).

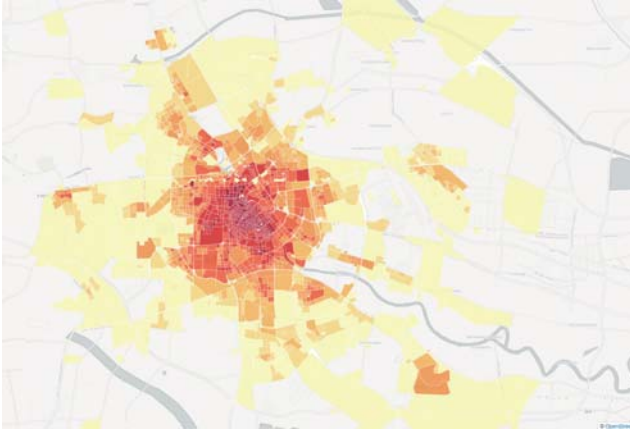| Models	Minkowski distance		
	8:00–10:00	16:00–18:00	20:00–22:00
Proposed model	0.604	0.753	0.787
Utility model	5.607	6.387	6.516
Potential model	2.315	2.972	2.744

Figure 8: Projected result of road condition in Tianjin (8:00 a.m.–10:00 a.m., weekend).

as $\left(\sum_{i=1}^{|V|} |m_i - y_i|^p\right)^{1/p}$, where m_i is the impedance of the block center v_i computed by each model, y_i is the road condition of block v_i, and p is set to 2. Here, a smaller distance means more similarity. Notice that the time factor is considered for all models. Table 1 shows the similarity between each model considered in this work and the actual road conditions.

Table 1 reveals that our proposed model yields results with the most similarity to actual road conditions. The utility model performs the poorest among the models studied, and the potential model shows moderate accuracy. The poor performance of the traditional models may be explained as follows: The utility model only considers the attractiveness of the destination and ignores the path between origin and destination, while the potential model takes the distance from the origin to the destination into consideration but ignores other parameters, such as turns and junctions. Our proposed model combines the parameters path length, number and angle of turns, number and direction of junctions, and population density, thereby accurately depicting the city's impedance at different time periods.

5. Conclusions

Urban impedance is an important indicator to consider in assessments of transportation and land-use systems. However, traditional impedance models require extensive data collection, which is costly, but yield only coarse-grained results. The present work leverages check-in records obtained from mobile social networks to build a fine-grained but inexpensive urban impedance model. We use check-in records

to adjust the path and population parameters of the model. Check-in records can filter functional locations and supply the time factor, which not only influences the function density but also impacts the real-time population of an area. Several experiments confirmed that our proposed model yields more accurate results than traditional impedance models.

In future research, we aim to take multiple-path factors into consideration and employ more types of data to obtain an improved impedance model.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] M. Batty, "Accessibility: in search of a unified theory," *Environment and Planning B: Planning and Design*, vol. 36, no. 2, pp. 191–194, 2009.

[2] N. He, N. Liu, and S. Zhao, "A study of road traffic impedance base on BPR function," *Journal of Nanjing Institute of Technology: Natural Science Edition*, vol. 11, no. 1, pp. 6–11, 2013.

[3] S. L. Handy and D. A. Niemeier, "Measuring accessibility: an exploration of issues and alternatives," *Environment and Planning A*, vol. 29, no. 7, pp. 1175–1194, 1997.

[4] M. Wachs and T. G. Kumagai, "Physical accessibility as a social indicator," *Socio-Economic Planning Sciences*, vol. 7, no. 5, pp. 437–456, 1973.

[5] W. G. Hansen, "How accessibility shapes land use," *Journal of the American Institute of Planners*, vol. 25, no. 2, pp. 73–76, 1959.

[6] D. Keeble, P. L. Owens, and C. Thompson, "Regional accessibility and economic potential in the European community," *Regional Studies*, vol. 16, no. 6, pp. 419–432, 1982.

[7] "Highway capacity manual," *Transportation Research Board*, vol. 12, no. 7, pp. 690–696, 1987.

[8] H. Spiess, "Conical volume-delay functions," *Transportation Science*, vol. 24, no. 2, pp. 153–158, 1990.

[9] W. Wang, M. Jing, and D. Liu, "Routing with capacity limitation and multi-paths based on LOGIT model and BPR impedance functions," *Journal of Highway and Transportation Research and Development*, vol. S1, pp. 81–85, 2012.

[10] K. B. Davidson, "The theoretical basis of a flow-travel time relationship for use in transportation planning," *Australian Road Research*, vol. 8, no. 1, pp. 32–35, 1978.

[11] Y. Wang, W. Zhou, and L. Lu, "Theory and application study of the road traffic impedance function," *Journal of Highway and Transportation Research and Development*, vol. 21, no. 9, pp. 82–85, 2004.

[12] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, article 38, 55 pages, 2014.

[13] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in *Proceedings of the 13th International Conference on*

Ubiquitous Computing (UbiComp '11), pp. 89–98, ACM, Beijing, China, September 2011.

[14] Y. Fu, H. Xiong, Y. Ge, Z. Yao, Y. Zheng, and Z.-H. Zhou, "Exploiting geographic dependencies for real estate appraisal: a mutual perspective of ranking and clustering," in *Proceedings of the 20th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 1047–1056, ACM, 2014.

[15] C. Chen, D. Zhang, B. Guo, X. Ma, G. Pan, and Z. Wu, "TripPlanner: personalized trip planning leveraging heterogeneous crowdsourced digital footprints," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1259–1273, 2015.

[16] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 151–158, 2016.

[17] J. Yuan, Y. Zheng, and X. Xie, "Segmentation of urban areas using road networks," MSR-TR-2012-65, 2012.

[18] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*, pp. 186–194, ACM, Beijing, China, August 2012.

[19] Baidu Map API, http://lbsyun.baidu.com.

[20] CartoDB, http://www.cartodb.com.

[21] Sina Weibo, http://weibo.com.

[22] Dianping, http://www.dianping.com.