Understanding how urban residents process road network information and conduct wayfinding is important for both individual travel and intelligent transportation. However, most existing research is limited to the heterogeneity of individuals’ expression and perception abilities, and the results based on small samples are weakly representative. This paper proposes a quantitative and population-based evaluation method of wayfinding performance on city-scale road networks based on massive trajectory data. It can accurately compute and visualize the magnitude and spatial distribution differences of drivers’ wayfinding performance levels, which is not achieved by conventional methods based on small samples. In addition, a systematic index set of road network features are constructed for correlation analysis. This is an improvement on the current research, which focuses on the influence of single factors. Finally, taking 20,000 taxi drivers in Beijing as a case study, experimental results show the following: (1) Taxi drivers’ wayfinding performances show a spatial pattern of a high level on arterial road networks and a low level on secondary networks, and they are spatially autocorrelated. (2) The correlation factors of taxi drivers’ wayfinding performances mainly include anchor point, road grade, road importance, road complexity, origin-destination length, and complexity, and each factor has a different influence. (3) The path complexity has a higher correlation with the wayfinding performance level than with the path distance. (4) There is a critical point in the taxi drivers’ wayfinding performances in terms of path distance. When the critical value is exceeded, it is difficult for a driver to find a good route based on personal cognition. This research can provide theoretical and technical support for intelligent driving and wayfinding research.

1. Introduction

Wayfinding was formally defined by Lynch as the consistent use and organization of sensory cues from the external environment [1–3]. Wayfinding is not randomly navigating in the environment but rather a purposeful activity from an origin to a destination by comprehensively mobilizing cognitive knowledge of the surroundings. It can be influenced by both environmental factors and individual differences [3–5]. For example, a person with high spatial cognitive ability will become lost in some places. Generally, there is more than one alternative path between a pair of origin-destination (OD) points in a road network. However, there exist optimal and poor paths considering the consumption of time or other limited resources (e.g., fuel, economic cost). The choice of different paths reflects each individual’s wayfinding performance level (WPL) [6]. In an urban environment, road networks, which are key to urban transportation systems, play an important role in the movement and circulation of people and resources, and they guarantee the vitality of a city. Investigating the differences in wayfinding performance of city residents on road networks and discovering wayfinding behavior patterns and their correlation factors are helpful for understanding the universal law of human spatial knowledge acquisition and spatial consciousness development and important for both individual travel optimization and intelligent transportation planning.
The quality of wayfinding is the integrated result of the individual features including age, sex, professional background, perceptual ability, spatial ability, and environmental features including size, symbols, and structure [7–9]. Much research has been undertaken in a broad range of disciplines, particularly the environmental, behavioral, and computer science fields, in examining the principles and factors related to wayfinding. From the perspective of observation means, they can be categorized into three types: based on questionnaire surveys, based on professional equipment observation, and based on location-aware data mining. Questionnaire surveys are mostly used in behavioral and cognitive science. Marie-Dominique and Patrick [10] took taxi drivers as the study objects and analyzed the differences in wayfinding performance and strategy differences between novice and skilled drivers by drawing routes on maps and time estimation. Through a questionnaire experiment, Gärling and Gärling [11] found that 69% of shoppers consciously plan their paths to shorten the overall journey, but whether they can find the shortest path depends on their wayfinding ability. Ishikawa and Montello [12] and Al-Azawi and Al-Alwan [13] explored the influence of repetition and familiarity on the level of wayfinding and spatial knowledge development from different angles based on hand sketching and questionnaires. Giudice et al. [14] and Kim et al. [15] focused on the effect of map presentation mode on pedestrian wayfinding. Giudice et al. studied the influence of map presentation mode on the level of environmental learning, cognitive map development, and wayfinding performance with visual impairment through controlled experiments. Kim et al. discussed the differences between digital signage and traditional signage in helping people find their way on a campus.

With the development of science and technology, the wide applications of virtual reality, eye movement recorder, video capture, real-time positioning, and other technologies provide more convenient and accurate observation methods for wayfinding behavior research [16–19]. Li and Klippel [20] used audio and video acquisition equipment to explore human wayfinding behavior in complex buildings and explored the interaction between individual differences and architectural features in the process of wayfinding. Wang et al. [21] discussed the effects of sex and age on wayfinding performance by using eye trackers. The results showed that there were significant interactive differences between sex and age in self-location memory and route memory. Malinowski and Gillespie [22] used positioning technology to carry out a wayfinding experiment of the field environment. It showed that the influence of individual differences found in the laboratory experiment, such as sex, previous experience, mathematical ability, and map skills, also exists in field tasks. In addition, the adoption of new technologies allows some experiments to be implemented in a laboratory. Laboratory-based research reduces the impact of the real, dynamic, and complex environment, which is convenient for variable control and detail discovery. Spiers and Maguire [23] used retrospective oral reporting protocols, eye tracking, and highly accurate virtual reality simulations of a real city (London, UK) to explore taxi drivers’ decision-making step by step. Taillade et al. [24] compared the wayfinding and spatial memory performance of young and elderly people by using virtual reality technology. It was found that age had a significant effect on wayfinding ability but had no significant effect on spatial memory ability. Brunye et al. [8] discussed the impact of time pressure on wayfinding in a virtual environment. The results showed that time pressure can increase the intuitive feeling of pressure and reduce the accurate performance of wayfinding tasks. Jansen-Osmann and Wiedenbauer [25] investigated the influence of structuring space on wayfinding performance, wayfinding strategies, and spatial knowledge in an unfamiliar virtual environment.

With the accumulation of location-aware data [26, 27], scholars also carry out wayfinding research through location-aware data mining. This type of research focuses on pattern discovery in path selection and simulation. Turner [28] proved that motorcyclists pay more attention to the smaller angle distance than to the shortest distance in a familiar environment. Tang et al. [29] and Li et al. [30] explored the route selection patterns of taxis and further applied them to path planning. Liu et al. [31] identified a set of valuable features through trajectory analysis to explore cab drivers’ operation patterns and compared cab drivers’ wayfinding behaviors. Zhang et al. [32] explored the travel mode and seasonal regularity of taxi groups, which provided useful information for trip generation.

The questionnaire survey methods have advantages in obtaining the attribute information and specific details of respondents but show strong individual heterogeneity because the retrospective description and question-and-answer format rely heavily on the memory, expression, and perception ability of respondents. Observation by professional equipment can obtain wayfinding process information comprehensively and accurately, which makes up for the drawbacks of questionnaire survey methods to a certain extent. However, this method can be applied only in experiments with small samples due to equipment limitations, causing weakly representative results. Although research based on location-aware data mining does not rely on prior knowledge, it mostly focuses on the discovery of path selection patterns but not on wayfinding performance. In summary, existing research mainly measures wayfinding behavior based on a small sample or in a small/indoor area and rarely explores wayfinding performance on city-scale road networks and underlying rules based on large sample data. For the correlation factor analysis of wayfinding performance, the existing studies focus more on a simple or single factor, such as color, repetition frequency, age, or sex, and few studies have paid attention to the influence of a comprehensive set of factors.

In view of the abovementioned problems, this paper aims at evaluating drivers’ wayfinding performances on road networks, exploring its spatial distribution characteristics and correlation factors based on trajectory data. The specific work and main contributions are as follows: (1) This paper proposes a quantitative and population-based evaluation method of wayfinding performance on city-scale road networks based on massive trajectory data. The method can
accurately compute and visualize the magnitude and spatial distribution differences of drivers' cognitive levels to the road network, which is not achieved by conventional methods based on small samples. (2) This paper constructs a systematic index set of road network features for correlation analysis of wayfinding performance. This is an improvement on the current research that focuses on the influence of single factors. (3) This paper discovers the spatial distribution characteristics and correlation factors of taxi drivers' wayfinding performances in Beijing.

The structure of this paper is as follows: Section 2 illustrates the study area and all datasets used in the experiment; Section 3 introduces the methodology for evaluating the wayfinding performance of drivers on a city-scale road network and analyzing its correlation factors; Section 4 presents the experimental results of the case study in Beijing; Section 5 discusses the implications and limitations of the study; and Section 6 concludes with future directions.

2. Study Area and Data

Beijing urban area is taken as the study area to investigate the spatial distribution and correlation factors of driver wayfinding performance, as shown in Figure 1. The data used in the experiment mainly comprise three datasets. The first dataset is taxi trajectory data from the taxi administration agency, and its attributes include the license plate number (encrypted for privacy protection), positioning time, passenger-carrying status, longitude, latitude, speed, and direction. Its sample dataset is shown in Table 1. The entire dataset is composed of approximately 350 million trajectory points collected from approximately 20,000 taxis from November 1 to 7, 2012. The taxi drivers are chosen as a study group for two reasons. First, taxi drivers generally have a larger scope of activity in a city and have more frequent interaction with the urban road networks than urban residents due to the nature of their work. Using a dataset of taxi drivers is conducive to analyzing wayfinding performance differences in the entire city. Second, it is difficult to obtain trajectory data of private cars, for the sake of privacy and security. The reason for selecting 2012 is that car-hailing and navigation applications were not used on a large scale by taxi drivers in Beijing at that time according to the development process of the car-hailing industry in China (https://www.didiglobal.com/about-special/milestone) and an interview survey of taxi drivers in Beijing, so taxi drivers’ wayfinding behaviors relied mainly on their cognition of road networks.

The second dataset is road network data, which include the road name, road grade, number of lanes, traffic direction, and longitude and latitude coordinates of the road nodes. The sample data are shown in Table 2. The main structure of the road network in Beijing is a combination of ring-shaped express roads and highways, and the secondary roads are evenly distributed across the city, as shown in Figure 2. The ring-shaped express roads mainly bear the traffic flow inside urban areas, while the highways bear the traffic flow between urban and suburban areas. The third dataset is a point-of-interest dataset (POI) including landmark buildings and anchor points. Its sample data are shown in Table 3. Both road network data and POI data in the year 2012 come from AutoNavi Maps, the largest map provider in China.

3. Methodology

Trajectory data provide an unprecedented and large amount of information that reflects the dynamics of mobile objects and thus are widely applicable to intelligent transportation, urban computing, social network analysis, and other fields [33, 34]. As far as this research is concerned, the trajectory data demonstrate the actual moving routes of taxis, while an objective optimal route based on road networks exists for each passenger-carrying trip. The discrepancy between them can reflect the wayfinding performance of taxi drivers. Based on this observation, a methodological flowchart is designed as shown in Figure 3, which is mainly composed of four parts: (1) In the data preprocessing stage, map matching between the original trajectory data and the road network is implemented. The driving route is reconstructed, and the passenger-carrying route segment is further extracted. (2) A quantitative evaluation index of wayfinding performance is defined as the WPL. The global index “global WPL” and local index “local WPL” are calculated based on the passenger-carrying route segments. (3) Based on the local WPL of all taxi trajectory data, the spatial distribution characteristics of the wayfinding performance are analyzed by the spatial autocorrelation analysis method. (4) The correlation factors and effects of the WPL are analyzed by statistical analysis methods from four aspects: feature point, road attribute, regional features of the road network, and OD features.

3.1. Data Preprocessing. The data preprocessing of this study mainly includes four parts:

(1) Map matching: The original trajectory data are the tracking points recorded by GPS receivers installed on taxis, and most of these points do not match road vectors due to positioning errors and signal receiving disturbances of GPS receivers. Therefore, the map matching must be conducted to match each trajectory point with the corresponding road segment. Considering that the map matching algorithm with a hidden Markov model has good performance [35, 36], we choose it for map matching in this study, and the accuracy of the map matching result is verified to be 87% using the accuracy ratio of points matched (ARP) from [37].

(2) Route reconstruction: For the consecutive trajectory points of each taxi, the matched road segments are connected in chronological order to reconstruct the complete driving path. In this study, the Dijkstra algorithm [38] is adopted to handle the matched road segments generated by sparse trajectory points. Moreover, in order to avoid the impact of related factors on the calculation result, the trajectories with a length of less than 1 km or with a dwelling time of more than 5 min in the course of the trip are removed before the index calculation.
(3) Passenger-carrying segment extraction: The taxis look for passengers in the cruising state, and the trajectories collected in this state reflect the cognition of the location of passenger sources rather than of road networks. Therefore, the route segments while carrying passengers are extracted according to the field value of the passenger-carrying state, and only these segments are used for further steps.

(4) Nighttime dataset extraction: This paper focuses on exploring the spatial distribution differences of taxi drivers’ WPLs and the correlation between WPL and static factors, so the impact of traffic flow on route selection is not considered at this stage. This study uses only the trajectory dataset collected in the nighttime (0:00–5:00), during which the traffic flow is relatively stable. In the end, 5 million positioning
data were screened out, covering the entire extent of the Beijing urban area.

3.2. Construction of Wayfinding Performance Level Index. Many studies have been carried out in the field of cognitive science and transportation for the aspect of wayfinding evaluation [23, 28, 39]. All scholars adopt some practical evaluation indexes from different perspectives, including the rate of wrong-way selections at intersections, the accuracy of map redrawing, the distance going the wrong way, the time spent viewing maps during wayfinding, and the time spent finding a destination. His study aims to explore the influence of the urban road network structure on a taxi driver’s wayfinding performance by using historical trajectory data of taxis. In the taxi industry, both the origin and destination become determined when a taxi picks up a passenger, and the taxi tends to reach the destination in the shortest path. The actual driving route can be regarded as the subjective optimal route based on the driver’s cognition on the road network, while there is an objective optimal route based on the road network. The discrepancy between the two routes can reflect the taxi driver’s wayfinding performance. Based on this observation, this paper chooses the ratio of the length of the shortest route to the actual driving route to construct the WPL [6, 9, 23] as follows:

\[
WPL = \frac{L_{\text{Shortest}}}{L_{\text{actual}}}
\]

where WPL is the wayfinding performance level index, \( L_{\text{actual}} \) is the length of the actual driving route, and \( L_{\text{Shortest}} \) is the length of the optimal route between the origin and destination of the trip by the Dijkstra algorithm based on the digital road network map. The greater the WPL value, the closer the actual driving route to the shortest route and the higher the WPL of the taxi driver.

3.3. Calculation of Wayfinding Performance Level Index. The driving route varies in length, and the WPL of the taxi driver on different road segments that the route passes by is also different. As shown in Figure 4, during the journey from Point A to Point B, the selected route is good overall, but it is not always a good choice in every segment (such as segment CD). Therefore, calculating only the overall wayfinding performance would ignore local detailed information and fail to reflect spatial differences.

Based on the above observation, this paper evaluates the driver’s wayfinding performance from the global and local perspectives, which correspond to global WPL and local WPL, respectively. The global WPL is calculated based on...
the entire driving route to evaluate the overall wayfinding performance on this route. A sliding window [40–42] with a width of 3 km and a step length of one road segment is designed to calculate the local WPL. As shown in Figure 4, the sliding window slides along the actual driving route (black solid line) progressively—e.g., from W1, W2 to W3 in the figure. For each window, the road network nodes at the two ends of the window are extracted as the start and end points, and the shortest path length between them is calculated based on road networks.

The WPL within the range of this sliding window (local WPL) is then calculated using (1). The result is assigned to the road segment within the current window to reflect the driver’s WPL to this segment. The local WPL is calculated in turn until all road segments along the driving route are processed. The wayfinding performance of a group of people to a road segment can reflect the difficulty of environment cognition in this region. If most drivers can make good wayfinding decisions in a certain region, there must be some features conducive to wayfinding in this area.

3.4. Spatial Pattern Analysis of Wayfinding Performance. Qualitative and quantitative methods are used to discover the spatial patterns that exist in the WPL of taxi drivers. Specifically, the global Moran’s [41] and local Moran’s [42] indexes are used to perform global and local spatial autocorrelation analysis on the WPL, respectively. The global spatial autocorrelation index is used to confirm whether there is a correlation between the WPL in a region and those in its neighboring region. By contrast, the local spatial autocorrelation index is used to explore the spatial location of the agglomeration center and corresponding patterns. The formula for the global Moran’s $I$ index is as follows:

$$I = \frac{n \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij})^2 (\sum_i (y_i - \bar{y})^2)}$$

(2)

where $I$ is the value of Moran’s $I$ index, $n$ is the total number of roads in the study area, $y_i$ and $y_j$ are the WPL values of roads $i$ and $j$, $\bar{y}$ is the mean WPL of all roads, and $w_{ij}$ is the spatial weight matrix between roads $i$ and $j$. 
The formula for local Moran’s I index is as follows:

\[ I_i = z_i \sum_{j \neq i} w_{ij} z_j, \]

where \( I_i \) is the value of the local Moran’s I index of road \( i \), \( z_i \) is the standardized value of the mean WPL, \( z_j \) is the standardized value of the standard deviation, and \( w_{ij} \) is the spatial weight matrix between roads \( i \) and \( j \).

The result of spatial autocorrelation analysis can help determine whether the WPL of taxi drivers is spatially autocorrelated and explore the degree of correlation and dependence. The two indexes are calculated by ArcGIS software, and the inverse distance weight method is used to construct the spatial weight matrix.

### 3.5. Potential Correlation Factors and Correlation Analysis

To the best of our knowledge, there is no previous literature specifically studying the influence of a city-scale road network on the wayfinding performance of drivers. In this paper, the selection of potential correlation factors mainly considers the comprehensive measurement of the structural characteristics of an urban road network. First, the urban road network is a network structure. Second, it has some unique features compared with common networks (anchor point, OD path). By referring to the literature on network structure characteristics [43, 44] and route planning on a road network [45, 46], this paper selects nine features as potential correlation factors: feature point, road grade, centrality (betweenness centrality, closeness centrality, straightness centrality), road density, road complexity, OD complexity, and OD distance. The details of each potential factor are shown in Table 4.

While calculating the indexes on centrality, this paper takes road intersections as nodes and road segments as edges to construct a graph of the road network. Then, taking multicenter analysis as a theoretical basis [43, 44], this paper uses the urban network analysis toolbox of ArcGIS software to calculate the multicenter measurement indexes.

In correlation analysis, we distinguish two types of variables: categorical variables and numerical variables. For categorical variables, the boxplot analysis method is used to display the distribution of WPL values across different attribute values, which shows the changing trends in WPL. For numerical variables, regression analysis is performed between them and WPL by Origin software to discover how WPL correlates with the factors.

### 4. Results and Analysis

#### 4.1. Spatial Distribution Characteristics of the Wayfinding Performance Level

According to the calculation method in Section 3.3, the average local WPL value of all taxis on the same road segment is calculated, and it reflects the WPL of the taxi drivers on this road segment. Figure 5 shows the spatial distribution of WPL. The high-WPL road segments marked by blue are mostly freeways, expressways, and major roads, which are basically consistent with the arterial roads in Beijing. Between the East 2nd Ring and the East 3rd Ring is a significant low-WPL area. Generally, the taxi drivers’ wayfinding performance in Beijing shows an obvious hierarchical pattern.

The global and local Moran’s I indexes are used to further analyze the clustering characteristics of WPL. Table 5 shows that the global Moran’s I is 0.2115, the \( z \)-value is 225.33, and the \( p \)-value is less than 0.05, which indicates that wayfinding performance has a strong spatial autocorrelation. Based on this finding, it can be inferred that if a driver can make a good wayfinding decision in a region, he or she is likely to have good performance in adjacent regions. Figure 6 shows the results of local spatial autocorrelation analysis, and most of the arterial roads show an obvious high-high aggregation pattern. There are several low-low aggregation areas distributed in the regions between the East 2nd Ring and East 3rd Ring, between the Northeast 4th Ring and Northeast 5th Ring, and between the West 2nd Ring and West 4th Ring. The high-low aggregation pattern is the phenomenon that highly cognized roads are surrounded by lowly cognized roads, which can be observed around the Northeast 2nd Ring, Northeast 3rd Ring, and Jinggang’ao freeway. The overall result is the following: (1) The wayfinding performance on road networks shows a spatial pattern of a high level on arterial road networks and low level on secondary networks. It is speculated that taxi drivers have a hierarchical wayfinding pattern from arterial networks to secondary networks that is step by step. (2) Wayfinding performance has a strong spatial autocorrelation. If a driver is familiar with a road, he or she tends to be familiar with adjacent roads. (3) There are several concentrated areas with weak wayfinding performance in road networks in Beijing in which it is difficult for most people to find an optimal way.

#### 4.2. Exploration on Correlation Factors of the Wayfinding Performance Level

##### 4.2.1. Feature Point

From the cognitive perspective, landmarks are important focal points and are prominent cognitive clues with distinctive features [47–50]. Based on the concept of landmark significance (i.e., semantic, visual, and structural significance), this study constructs a feature point dataset as shown in Figure 7, which is composed of landmark buildings (e.g., Tiananmen Square and National Stadium), road anchor points (e.g., overpasses, underpass tunnels, and roundabouts), and others [51, 52]. Figure 8 shows the average WPL of all road segments within 500 m of the feature points, road anchor points, and landmark buildings, and the red line indicates the average value of WPL for all road segments. Generally, the taxi drivers’ WPLs near feature points are higher than average, among which the WPL around road anchor points is more prominent, far beyond the average value. It can be speculated that taxi drivers in Beijing take road anchor points as important marks for cognizing road networks, and these anchor points play a greater role in route planning.

##### 4.2.2. Road Attributes

Road attributes reflect the role and function of a road in a road network system, and two attributes (road grade and centrality) are selected for analysis. In Figure 9,
### Table 4: Definition and expression of potential correlation factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature point</td>
<td>Road anchor points (overpass, highway entrance and exit, bridges) and landmark buildings.</td>
<td>Feature points are defined as important focal points and cognitively salient cues with prominent features.</td>
</tr>
<tr>
<td>Road grade</td>
<td>Freeway, expressway, trunk road, secondary trunk road, branch way.</td>
<td>It identifies the grade, functions, and traffic volume of the road.</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>$C_{AB} = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{jk}(i)$</td>
<td>It indicates the force and influence of nodes in the whole traffic network. The greater the betweenness centrality of nodes is, the more the nodes can control the relationship between other nodes.</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>$C_{API} = 1/\sum_{i=1}^{n} d_{ij}$</td>
<td>It indicates the difficulty for a node to reach other nodes, and the greater the closeness centrality, the better the accessibility of the node.</td>
</tr>
<tr>
<td>Straightness centrality</td>
<td>$C_{AS} = 1/(N-1)\sum_{i,j \neq k}^{N} (d_{ij}^{Eucl}/d_{ij})$</td>
<td>It indicates the accessibility efficiency between one node and other nodes. The greater the value of this index, the closer the actual route to a spatial straight line.</td>
</tr>
<tr>
<td>Road density</td>
<td>$D_i = R_i/S_i$</td>
<td>The ratio of the total length of the roads to the total area (km/km²).</td>
</tr>
<tr>
<td>Road complexity</td>
<td>$C_i = N_i/S_i$</td>
<td>The ratio of the number of road intersections to the total area (/km²).</td>
</tr>
<tr>
<td>OD complexity</td>
<td>The number of road intersections passed by the shortest route between the origin and destination.</td>
<td>It indicates the number of turning choices between the origin and destination.</td>
</tr>
<tr>
<td>OD distance</td>
<td>The length of the shortest route between the origin and destination.</td>
<td>It indicates the total driving distance.</td>
</tr>
</tbody>
</table>

*Figure 5: Spatial distribution of taxi drivers’ local wayfinding performance levels.*

The table above presents various factors and their expressions that are used to assess the potential correlation factors in the context of local wayfinding performance levels. Each factor is defined with a specific expression, and the description column provides a detailed explanation of the significance and impact of each factor on the overall performance.
the boxplot describes the statistical characteristics of the wayfinding performance for each road grade, and the red line represents the average WPL of all roads. The wayfinding performance on a freeway, expressway, and trunk road is higher than the average value. With a decrement in the road grade, the average WPL decreases, and the fluctuation amplitude of WPL increases. From the perspective of the application, a perfect and appropriate design of an arterial urban road network is very important and provides urban residents with a more friendly cognitive experience.

Figures 10(a)–10(c) show the distribution of the betweenness, straightness, and closeness centrality of the road network in Beijing. The distribution structure of the betweenness and straightness centrality is polycentric, while the closeness centrality shows a single-center distribution and conforms to the distance attenuation law from the center to the outside. To explore the relationship between the road centrality and WPL, this study shows a changing trend in WPL with the centrality index by a scatterplot, as shown in Figure 11. The dotted line is the fitting line of the scatterplot, while the broken line demonstrates the change rate of WPL with the centrality index.

Figures 11(a)–11(c) show that the WPL and the three types of centralities are positively linearly correlated, second-order correlated, and negatively linearly correlated, respectively. Note that the sparse and noisy points in each figure can be neglected since they are extraneous samples, which cannot reflect the general trend. The betweenness centrality indicates the function of transfer and switch of a segment in a road network, which corresponds to the importance of a road to some extent. As shown in Figure 11(a), the higher the betweenness centrality, the more important the road segment in the entire network and the higher the WPL among taxi drivers. The straightness centrality measures the degree of deviation between the shortest route and the straight path between two nodes, which indicates the simplicity of a route. As shown in Figure 11(b), the higher the straightness centrality, the higher the WPL among taxi drivers. In addition, the wayfinding performance is very sensitive to the straightness centrality, so their correlation is a second-order function. The closeness centrality is how close a node is to other nodes in the road network, which reflects the alternative road choices available in the neighboring area in route selection. As shown in Figure 11(c), the greater the closeness centrality is, the more difficult the road segment is cognized and the lower the WPL among taxi drivers is. In the future, a comprehensive consideration of the road capacity and cognition in urban road network planning will help build a more public-friendly road network.

<table>
<thead>
<tr>
<th>Moran’s I</th>
<th>p-value</th>
<th>z-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2115</td>
<td>&lt;0.05</td>
<td>225.33</td>
</tr>
</tbody>
</table>

Figure 6: Local spatial autocorrelation analysis of taxi drivers’ local wayfinding performance levels.
4.2.3. Regional Features of Road Networks. The study area is divided into grids with a size of 1 km\(^2\) 1 km, and then the average local WPL, road complexity, and density within each grid are calculated, as shown in Figure 12(a)–12(c), respectively. The areas with high road complexity are distributed mainly between the 2nd Ring and 4th Ring, especially along ring expressways (Figure 12(b)). The distribution of road density is similar to that of road complexity. The area within the 4th Ring and the area north of the 5th Ring have a significantly high road density (Figure 12(c)). As shown in Table 6, the correlation analysis results indicate that the WPL of the regional road network is obviously correlated with road density and complexity. The correlation between the road complexity and WPL (\(R^2\) is 0.90) is much stronger than that between road density and WPL.

4.2.4. OD Features. To reveal the relationship between the OD features and WPL, this study takes the path complexity and distance as the independent variables and the global WPL as the dependent variable to perform linear regression analysis, as shown in Figure 13. Global WPL has a negative correlation with the number of road intersections passed by...
Figure 10: Continued.
Figure 10: Road centrality distribution: (a) betweenness centrality; (b) straightness centrality; (c) closeness centrality.

Figure 11: Continued.
Figure 11: Correlation analysis of road centrality and local WPL: (a) betweenness centrality; (b) straightness centrality; (c) closeness centrality.

Figure 12: Continued.
It can be seen that 140 is a critical point before which the linear relationship between the two variables becomes more significant ($R^2$ is 0.9063, as shown in the small subfigure of Figure 13(a)), after which the decreasing trend of global WPL fluctuates greatly as the path complexity grows.

H—there is a similar correlation pattern between global WPL and path distance. Global WPL has a negative correlation with the path distance ($R^2$ is 0.3848). Here, 20 km is a critical point, before which the linear relationship is more significant ($R^2$ is 0.7881, as shown in the small subfigure of Figure 13(b)), after which the correlation becomes insignificant.

Generally, it is speculated that there is a critical point in the taxi drivers’ wayfinding performances on road networks in terms of path distance. When the critical value is exceeded, it seems that it is difficult for a driver to find a good way based on personal cognition. Compared with path distance, path complexity has a higher correlation with WPL. A driver who encounters a road intersection must make a route choice, so the number of intersections passed by can better reflect the difficulty of driving and navigation. It can be inferred that the number of road intersections has a greater impact on the difficulty of wayfinding than on the driving time.

<table>
<thead>
<tr>
<th>Table 6: Correlation between local WPL and road complexity and density.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Road complexity</strong></td>
</tr>
<tr>
<td>R squared coefficient</td>
</tr>
</tbody>
</table>

\( y = -0.0011x + 0.9092 \quad R^2 = 0.9063 \)

\( y = -0.0006x + 0.8752 \quad R^2 = 0.8088 \)

\( y = -0.0023x + 0.8446 \quad R^2 = 0.7881 \)

\( y = -0.0034x + 0.8585 \quad R^2 = 0.8464 \)

\( y = -0.0003x + 0.8305 \quad R^2 = 0.3848 \)

Figure 12: Characteristics of the regional road networks: (a) local wayfinding performance level; (b) road complexity; and (c) road density.

Figure 13: Correlation analysis of OD features and WPL: (a) path complexity; (b) path distance.
5. Discussion

The spatial distribution map of WPL shows that taxi drivers have a good understanding of the skeleton of a road network, even in the Central Business District of Eastern Beijing, which has a complex network structure. It can be inferred that drivers’ cognition of the urban road network may have such a hierarchical cognitive mode from the spatial distribution of wayfinding performance, which makes the driver’s wayfinding performance gradually decrease from the arterial road network to the secondary road network, but further study of this dynamic cognitive process is still required.

Experimental results show that feature points and road grade are important references in route selection and play an important role in wayfinding, and among all types of feature points, taxi drivers have a higher WPL to road anchor points (i.e., overpasses and roundabouts). In addition, this paper explores the relationship between road centrality and WPL for the first time. The results show that there is a close and predictable relationship between road centrality and wayfinding performance. Road centrality does not exist explicitly in real space as a concept of graph theory, but it is still closely correlated to wayfinding performance.

The results in Section 4.2.4 prove that as the path distance increases, the global WPL decreases, which means that the deviation of the actual driving route from the shortest route shows a cumulative growth trend, and the deviation value can be predicted to some extent. However, the deviation becomes unpredictable when the driving route exceeds 20 km or the number of passed by intersections exceeds 140. It can be inferred that there might be an upper limit for the human brain to cognize a road network. When the upper limit is exceeded, it is difficult to plan a good route based on personal cognition, and the route selection results are relatively random. It is necessary to further confirm the reasons causing a long route planning failure [50].

At present, most wayfinding performance research takes a trajectory as the evaluation unit [23, 28, 39], but taking only the global wayfinding performance result as a performance indicator is not a good choice because the driver does not have the same level of cognition on every segment along a long route. This study applies a sliding window to calculate the local wayfinding performance (local WPL) to reflect the cognitive status of drivers on different road segments along a route. Evaluating the global and local WPL at the same time can help reveal more fine-grained spatial wayfinding decision patterns.

This research provides insights into drivers’ perceptions of road network characteristics during wayfinding, which is an important research direction in cognitive science, transportation, and geography [53]. It has high theoretical value: (1) It can prepare the ground for the simulation and forecast of travelers’ behaviors under hypothetical scenarios and further close the gap between urban features used by drivers during route choices and the computational representation of these features used in modeling this process. (2) For driverless vehicles, it can provide a theoretical basis for the autonomous learning and brain-like decision-making of vehicles. (3) With the development of indoor positioning technology (gyroscope positioning, WiFi positioning, video positioning), it has become possible to track the continuous trajectory of pedestrians in various environments, for example, within buildings. Therefore, the proposed research framework can also aid in investigating the wayfinding performance of pedestrians in these environments. The population-based experiment will promote the understanding of pedestrians’ wayfinding behaviors in an indoor environment. In addition, this research has a number of practical application values: (1) Existing navigation software conducts route planning based on limited deterministic information but cannot consider certain complicated information. Integrating the group wayfinding experience into route planning is conducive to improving the accuracy of navigation. (2) The research results can help improve the information presentation mode and user experience of navigation products. For example, instructed by the research results, the expression of route planning can be divided into high-grade and low-grade road parts to improve understandability. The former part takes road anchor points as the description unit (e.g., “Go straight to Madian Bridge, turn right onto the North Third Ring”), while the latter takes the distance as the description unit (e.g., “Go straight for 7 km, turn left”).

However, more research questions should be further investigated to enhance this work. Individual heterogeneity is an indispensable point in the study of wayfinding [47], but this paper focuses on the common wayfinding characteristics of the entire population of taxi drivers instead of individuals. Moreover, this study explores the correlation factors of taxi drivers’ WPLs separately. Nevertheless, each factor is analyzed separately while conducting correlation analysis at this stage, so it is necessary to further consider the comprehensive influence of road network characteristics and explore the leading factors. In addition, with the popularity of navigation and car-hailing applications, the locality and uncertainty of taxi drivers’ cognition of road networks gradually decrease. Therefore, exploring the changes in the wayfinding modes to road networks in this period of new technology is important for understanding the impact of auto-navigation on human spatial cognitive processes. Third, human activities rely on the physical environment and bring new meanings to the physical environment (i.e., context information). This paper currently focuses on the features of the pure physical environment (e.g., road density) while investigating the correlation factors of wayfinding performance and has not yet considered the features of context information (e.g., traffic flow caused by human travel, and type of trip). It may be interesting to study wayfinding based on contextual information in the future. More comprehensive research requires the interdisciplinary fusion of psychology, physiology, behavioral science, geography, computer science, environmental science, and other disciplines.

6. Conclusions

This paper proposes a quantitative and population-based evaluation method of WPL based on massive trajectory data. It can accurately compute and visualize the magnitude and spatial...
distribution differences in drivers’ wayfinding performances, which is not achieved by conventional methods based on small samples. In addition, a systematic index set of road network features is constructed for correlation analysis of wayfinding performance, including point features, regional features, attribute features, and OD features. Finally, taking the taxi drivers in Beijing in the year 2012 as a case study, we analyzed the spatial distribution characteristics of taxi drivers’ WPLs and the correlation factors. Experimental results quantitatively reveal that the wayfinding performance is hierarchically distributed and spatially autocorrelated, and its correlation factors mainly include anchor features, road grades, road importance, road complexity, OD length, and complexity. This research is useful for understanding people’s wayfinding performance characteristics on road networks and provides theoretical and technical support for intelligent driving and wayfinding research. The large-scale study based on trajectory data lacks the attribute information of individual participants such as sex, age, and occupation, so it cannot reflect the influence of individual attributes on wayfinding performance. Future studies will be conducted to deepen the understanding of wayfinding performance on city-scale road networks by considering individual heterogeneity, joint influence of various features, and context information (e.g., traffic flow caused by human travel, type of trip).

Data Availability
The data used to support the findings of this study came from the taxi management agency of Beijing.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Authors’ Contributions
Jun Li contributed to conceptualization and funding acquisition; Zhenwei Li performed data curation; Yan Zhu and Xiao Sang carried out formal analysis; Jun Li and Yan Zhu participated in methodology and wrote the original draft; Zhenwei Li and Wenle Lu were responsible for software; Wenle Lu created visualization; Yang Ji and Xiao Sang wrote, reviewed, and edited the paper.

Acknowledgments
This research was financially supported by the National Natural Science Foundation of China (No. 41971355), the Open Project of State Key Laboratory of Resources and Environmental Information System, and the Yueqi Young Scholars Program of China University of Mining and Technology at Beijing.

References


