Research Article

Identification of Working Trucks and Critical Path Nodes for Construction Waste Transportation Based on Electric Waybills: A Case Study of Shenzhen, China

Jun Bi,1,2 Qiuyue Sai,2 Fujun Wang,2 and Yakun Chen2

1Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China
2School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Qiuyue Sai; 19114026@bjtu.edu.cn

Received 5 January 2022; Accepted 25 February 2022; Published 22 March 2022

Academic Editor: Sang-Bing Tsai

Copyright © 2022 Jun Bi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the large amount of waste generated by urban construction, the transportation of construction waste has a significant impact on urban traffic. Understanding the transportation trajectory of garbage trucks can improve the management of transportation routes and reduce traffic accidents. This study analyzes electric waybill and state data of garbage trucks to identify hot nodes of construction waste transportation, where the volume of garbage trucks is relatively high. Management should strengthen the hot nodes to reduce traffic accidents. First, several machine learning methods are used to improve the prediction accuracy of electric waybill generation, where the garbage truck recorded on the electric waybill is regarded as a working truck. Second, the transportation trajectory of working trucks is extracted, and its spatiotemporal characteristics are further analyzed. Hot nodes are found based on density clustering. Finally, a case study is conducted based on the Shenzhen construction waste transportation system. The results show that the XGBoost model can improve the accuracy of the generation of waybill to 90.5% compared with the decision tree model, random forest, and GBDT. Moreover, the density clustering model can discover the hot nodes of construction waste transportation. Considering the minimum number of samples and the neighborhood radius, the clustering number is determined as 100. The ratio of noise points is determined as 0.79. The results can provide decision support for the management of electronic waybill and garbage truck transportation.

1. Introduction

Along with urbanization, the renovation and expansion of construction have produced a large amount of construction waste, which has put pressure on the urban environment and traffic [1]. Construction and demolition waste (CDW) is one of the largest waste streams and needs to be transported by cities [2]. Serious traffic accidents can easily occur because of working garbage trucks. Working garbage truck is a garbage truck that is carrying out the transportation task, which has a large weight and size. Working garbage truck has large inertia and large blind spot and is difficult to control. The road it passes is often the key safety management object. Therefore, it is important to analyze the transport trajectory of working garbage trucks to strengthen transportation management [3]. Identifying working garbage trucks is important. Whether the trucks carry out the transportation task is judged according to the waybill information. If a truck is carrying out a task, it is recorded by a waybill. Historically, a paper waybill was used to record the transporting process. However, it requires considerable manpower to record information, which causes the problem of incomplete supervision. To manage the process of construction waste transportation efficiently, electronic waybills have been used in waste transportation based on technologies such as communication technology and the Internet of Vehicles. Electronic waybills are able to store transportation information in the system in the form of electronic data. Applying an electronic waybill can save manpower and record the transportation process online. The state data of garbage
trucks on the construction sites are used to judge the waybill generation. Only the garbage truck carrying out a waybill can be defined as a working garbage truck that is carrying out the transportation task. However, an electronic waybill is automatically generated based on the GPS of vehicles. Because the state of garbage trucks is not considered, a waybill is generated by mistake if the garbage truck passes the construction site without a transport mission, which results in the low prediction accuracy of waybill generation in general and thus contributes to the low identification accuracy of working trucks. Therefore, an accurate prediction model for electronic waybill generation is urgently needed.

After the working trucks are identified, their transport trajectory can be analyzed. Due to the long transportation path from construction sites to suburbs, vehicle management cannot be comprehensive. In general, the transportation of construction waste has a higher risk for traffic accidents because the volume of garbage trucks is high, which may cause heavy losses to transport contractors and society. An effective method for avoiding traffic accidents during construction waste transportation is to analyze the nodes on the road where the accident may occur [4]. Specifically, the transport trajectory can be explored to find the important nodes.

Due to the acceleration of urbanization and the emergence of new technologies, the management of construction waste transportation routes is a new problem. The urban construction waste production of China has increased. Construction waste transportation has put enormous pressure on urban traffic, so the management of construction waste transportation is very urgent. We reviewed a lot of literature on construction waste research. Several studies have discussed construction waste transportation from different aspects. For instance, some studies have focused on the impact of the recycling of construction waste on the environment. Lachat et al. [5] presented a life cycle inventory compilation and life cycle assessment in France. Souza et al. [6] analyzed the impact of recycling proposals for construction waste on the environment. Maués et al. [7] evaluated the environmental impact generated by transportation construction waste. The results showed that transportation waste management needs to be strengthened to improve the sustainability of cities. In addition, studies have focused on construction waste management. Tao and Xiao [2] analyzed the quantification and composition of construction waste in Shanghai, China, and recycling management in this region was discussed and introduced. Franco et al. [8] applied a model to optimize the location of landfills, and it considered the cost of transporting waste and the cost of building the landfill. Spišáková et al. [9] confirmed the economic potential of CDW audit processing, and disposal costs and transport costs of the recommended CDW management were considered.

Moreover, studies have also concentrated on developing construction waste management system design. You et al. [10] proposed an informatization scheme integrating multiple technologies, which was used to monitor illegal behaviors in the waste disposal process. Wang et al. [11] combined building information modeling technology (BIM) and vehicle positioning technology (GIS) to develop a monitoring and intelligent management information management platform for construction waste, which improved the precision and intelligence of construction waste management. Wang et al. [12] developed a BIM and cost-optimization-based decision-making system for construction waste transportation that suggested a cost-effective transportation plan. Zhang [13] took underground engineering construction as the research object and analyzed the source characteristics of construction waste. To resolve the disadvantages of construction waste management, construction waste construction site management strategies have been proposed to improve the resource utilization of construction waste. In [14], an intelligent urban construction site muck monitoring system was developed and the system was deployed to the cloud server combined with vehicle GPS positioning and video remote monitoring. It improved construction waste transportation management under the data integration and network integration environment. Previous studies on construction waste management mainly focused on the environmental impact, resource utilization, and system design. However, few studies have focused on the characteristics of construction waste transportation trajectories.

Among the researches on transportation problems, some studies focus on the use of electronic waybills in transportation systems. Bakhtryar et al. [15] analyzed information synergy between e-Waybill solutions and intelligent transport systems. Cane et al. [16] designed an electronic multimodal waybill and a solution for implementation using the e-Freight e-Delivery Infrastructure. However, there are few studies on how to apply electronic waybills in construction waste transportation.

In addition, with regard to the study on transportation routes, some studies focused on traffic problems [17, 18]. Machine learning methods can be used to learn from large volumes of data [19, 20]. Some studies found traffic hotspots using clustering methods. Ran et al. [21] introduced a novel K-means clustering algorithm based on a noise algorithm to capture urban hotspots. Jia et al. [22] analyzed traffic crash with point-of-interest spatial clustering. Le et al. [23] found traffic accident hotspot based on kernel density estimation. DBN clustering method is used for key node identification of construction waste transportation path. The use of electronic waybill data and route trajectory data to analyze important nodes in the transportation route of construction waste fills the research gap.

The study aims at solving the problem of frequent traffic accidents in construction waste transportation. The innovative idea is to transform the transport routing problem into a node management problem. These nodes need to be managed centrally. First, it is important to find the trajectory of the trucks at work. The working status of the truck is recorded in the electronic waybill [24]. Based on these data, it is predicted whether the trucks are under working states. XGBoost is designed to judge the working state of the trucks. XGBoost used boosting method to improve the prediction efficiency and accuracy, which is widely adapted to build a prediction model. Second, for trucks in the working state,
the density clustering method is used to find the concentration of trucks at a certain time. The clustering method is to classify based on the density of spatial distribution. Areas where the working truck congregate will be classified into one category [25]. This is the area in the transportation route that should be concentrated.

The contribution of this study is shown as follows. The study aims at solving the problem of frequent traffic accidents in construction waste transportation. We proposed an innovative idea that transforms the transport routing problem into a node management problem. We have found the nodes that need to be managed centrally. Strengthening the management of important nodes can improve management efficiency and reduce the probability of traffic accidents. Moreover, we obtained a model with better accuracy through case studies among several methods. Potential applications of the proposed method are shown as follows. First, based on the emerging electronic waybill and electronic fence technology, machine learning technology is used to design the identification method of working trucks and key nodes in this study. The key node is the area with frequent traffic accidents, which can make the loss of life and property. Therefore, it is helpful to find important path nodes in transportation routes and strengthen management to ensure a healthy transportation environment. Second, electronic fences and electronic waybills are new types of applications. Replacing paper waybills with electronic waybills can save 65%–99% of time [15]. However, the application of electronic waybill technology in the construction waste transportation process is rarely applied. This study can expand the application of electronic waybill technology. The proposed methods may be used by transport contractors and related stakeholders to manage construction waste transport roads.

The remainder of this article is arranged as follows. Section 1 introduces the problem background and main contributions in this study. Section 2 introduces data and data preparation and addresses the problem. Section 3 explains the methodology used to solve the problem. Section 4 presents a case study on Shenzhen and provides managerial insights. Finally, Section 5 provides the conclusions and suggestions for future research.

2. Data Preparation and Definition

2.1. Area and Data Introduction

2.1.1. Study Area Introduction. Construction waste transportation in Shenzhen, China, is explored in this study. Based on the GPS system [26], an electronic waybill is used to manage the whole process of construction waste, including generation, transportation, and disposal. Shenzhen has more than 9084 garbage trucks and 4000 construction sites. The distribution of construction sites in Shenzhen is shown in Figure 1, where the red points represent the locations of construction sites.

2.1.2. Data Introduction. The trajectory data of the garbage truck are obtained by a GPS positioning device. Trajectory data comprise on-road and waybill node information. A total of 128.7 million GPS trajectory data points from 9084 garbage trucks are obtained. Track data fields include track ID (Track_ID), speed, vehicle number (Vehicle_ID), longitude of truck (LNG), latitude of truck (LAT), mileage (Mil), and track GPS data time (GPS_Time). The sample data are shown in Table 1.

A total of 13,947 pieces of state data of vehicles in the waybill nodes are obtained. The data are used for the prediction model of waybill generation. The aim is to identify whether to generate an electronic waybill. The sample data are shown in Table 2.

2.2. Construction Waste Transportation and Node Definition. A waybill is used to record the complete construction waste transportation process. The process is shown in Figure 2. When the garbage truck enters the electronic fence of the first construction site, a new waybill is generated. Then, when it passes through a construction site, a waybill node corresponding to the construction site is generated. A waybill can have several waybill nodes if the garbage truck needs to carry out transportation missions at several construction sites. After the truck drives into the electronic fence of the disposal site, the end node is generated and the waybill is finished. Some definitions are as follows:

1. Electronic fences indicate areas for construction. The truck enters and exits the area for related processing procedures. Electronic fences include construction sites and disposal sites, and they are marked with the red square in Figure 2.

2. Electronic waybills record transportation process, including transportation from the source of the construction site to disposal at the end of the disposal site. A disposal site and more than one construction site are included in one complete waybill. If the system identifies that the garbage truck enters a waybill node to work, it will push the electronic waybill to the staff to confirm.

3. A waybill node in the waybill is generated at a construction site. If the garbage truck carries out a task in the electronic fence, a waybill node will be recorded on the electronic waybill.

4. The path node is the hot node on the transportation path, which is the blue circle marked in Figure 2. The volume of garbage truck at hot node is large.
3. Methodology

3.1. Waybill Node Generation Prediction Based on XGBoost.

Previous waybill generation prediction methods were judged using trajectory data, which usually led to misjudgment and low prediction accuracy. To improve the waybill generation prediction accuracy, state data of garbage trucks are considered here to identify waybill generation. The relationship between waybill generation and vehicle state data is complicated. Therefore, a decision tree model can be constructed to improve the waybill generation prediction accuracy based on the simultaneous use of trajectory data and state data of garbage trucks.

Improved decision tree models, such as random forest, GBDT, and XGBoost, in which multiple basic trees are combined, have been constructed [27–30]. Improved decision tree models can enhance computational efficiency and adaptability for prediction operations with large-scale datasets compared to basic decision tree models. In this study, four decision tree models are constructed to fit the prediction model of an electronic waybill. XGBoost is one of the most efficient decision tree models [31]. The calculation process of XGBoost is as follows [32].

First, the iteration of the objective function and Taylor expansion is shown in equation (1). The loss function is shown as equation (2).

\[
\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F, \quad \text{(1)}
\]

\[
\text{Obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \quad \text{(2)}
\]

where \(k\) is the number of basic decision tree models; \(f_k(x_i)\) is the \(k\)-th decision tree model; \(\hat{y}_i\) is the prediction result of garbage truck data \(x_i\) obtained by integrating \(K\) decision tree models; \(l(y_i, \hat{y}_i)\) is the training error; \(\Omega(f_k)\) is the regularization of the \(k\)-th tree; \(\Omega(f_k) = \gamma T + \frac{1}{2\lambda}||w||^2\); \(T\) is the number of leaf nodes; \(w\) is the value of leaf nodes; \(\gamma\) is the...
regularization coefficient of L1; λ is the regularization coefficient of L2. The optimal parameters and optimal model \( f^* (x) \) are obtained to minimize \( \Omega (f_k) \).

However, it is difficult to calculate the optimal model. Therefore, the problem is transformed to find the weight \( w \) and the structure of trees [33, 34]. The initialization model has no tree model. The prediction result is 0. Add the \( t \)-th tree to the model as shown in the following:

\[
\tilde{y}_i^{(t)} = \sum_{k=1}^{t} f_k (x_i) = \tilde{y}_i^{(t-1)} + f_t (x_i). \tag{3}
\]

The Taylor expansion is used to approximate the objective function, as shown in the following equation:

\[
\text{obj}^j (\theta) = \sum_{j=1}^{y} \left[ g_i f_r (x_i) + \frac{1}{2} h_i f_r (x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2. \tag{4}
\]

Take the derivative and obtain the optimal result, as shown in the following equations:

\[
w_j^* = \frac{\sum g_i}{\sum h_i + \lambda} \tag{5}
\]

\[
\text{obj}^* = \frac{1}{2} \sum_{j=1}^{T} \left( \frac{\sum g_i}{\sum h_i + \lambda} \right)^2 + \gamma T. \tag{6}
\]

3.2. Hot Nodes Identification Based on DBSCAN. Hot nodes are areas where the volume of garbage trucks is high. Serious traffic accidents often occur in areas with a high volume of garbage trucks. Thus, identifying hot spots and strengthening management can help alleviate traffic problems. In this study, the number of clusters for hot node recognition is not known in advance. The number of clusters is determined according to the hot node aggregation density. Therefore, DBSCAN is used to study the hot node recognition of construction waste transportation [35].

Two important parameters are used to describe the sample distribution, the radius of neighborhood \( \varepsilon \), and core point threshold MinPts. \( X = \{x_1, x_2, \ldots, x_n\} \) is the trajectory data of garbage trucks. For \( x_i \in X \), its neighborhood \( \varepsilon \) includes data defined as \( N_\varepsilon (x_i) = \{x_j \in X \mid \text{dist}(x_i, x_j) \leq \varepsilon \} \). For \( x_j \in X \), its density is \( \rho (x_j) = |N_\varepsilon (x_j)| \). For \( x_j \in X \), if \( N_\varepsilon (x_j) \geq \text{MinPts} \), then \( x_j \) is a core point. The collection of the core point is \( X_c \) [36].

The trajectory data of 200 garbage trucks during the evening peak of 20:00–21:00 are selected as an example. The sample size is 33,000 trips with latitude and longitude. This is a spherical surface; therefore, the latitude and longitude data should be converted to the actual distance in the rectangular coordinate system. Assume that the latitude and longitude of the two points are \((x_1, y_1)\) and \((x_2, y_2)\). The radius of the Earth is 6371 kilometers. The actual distance between the two points is shown in the following equation [37]:

\[
dist = \text{Rarccos} [\cos (x_1 - x_2) \cos y_1 \cos y_2 + \sin y_1 \sin y_2]. \tag{7}
\]

The calculation process is shown in Table 3 [37].

4. Case Study

4.1. Case Study on Waybill Node Generation Prediction. The scenario of the case study is introduced in Section 2.1, where construction waste transportation is explored in Shenzhen, China. The waybill is identified and generated in the waybill node based on GPS. To solve the problem of poor accuracy of the current electronic waybill identification, the characteristics of the garbage truck state data in the waybill node are analyzed. The generation prediction model of the electronic waybill at the construction site is constructed based on improved decision tree methods. The aim is to provide a high-accuracy prediction model to improve the quality and efficiency of construction waste transportation. First, the correlation of the influencing factors is analyzed. Then, the decision tree, random forest, GBDT, and XGBoost methods are used to fit the prediction model. Finally, the prediction results are compared and analyzed.

4.1.1. Variable Independence Test. The types of influencing variables and the objective (whether to generate an electronic waybill) are categorical data. Analyzing whether there is a dependency between two categorical variables is called an independence test. \( \chi^2 \) is used to perform statistical analysis on the correlation between each influencing variable and the objective to determine the input variables of the waybill generation prediction model. The main process is as follows:

1. The original hypothesis is \( H_0 \): there is no dependency between the two categorical variables. \( H_1 \): there is a dependency between the two categorical variables.
2. Calculate the expected frequency value, as shown in the following equation:

\[
f_{ij}^e = \frac{\sum_{i=1}^{R} \sum_{j=1}^{C} f_{ij}}{n} \times \frac{\sum_{i=1}^{R} f_{ij}}{n} \times \frac{\sum_{j=1}^{C} f_{ij}}{n}, \tag{8}
\]

where \( f_{ij} \) is the actual frequency of variables \( i \) and \( j \); \( f_{ij}^e \) is the expected value of the frequency; \( \sum_{i=1}^{R} f_{ij} \) is the frequency of variable \( i \); \( \sum_{j=1}^{C} f_{ij} \) is the frequency of variable \( j \); \( n \) is the sample size; \( R \) is the number of variable \( i \); \( C \) is the number of variable \( j \).
3. Calculate the statistics \( \chi^2 \) and degrees of freedom \( df \) as shown in the following equations:

\[
\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(f_{ij} - f_{ij}^e)^2}{f_{ij}}, \tag{9}
\]

\[
df = (R - 1) (C - 1). \tag{10}
\]

4. Given the significance level \( \alpha = 0.5 \), according to the calculated statistics \( \chi^2 \) and degrees of freedom \( df \), find the Chi-square distribution to obtain the value \( P \). If \( P > \alpha \), then \( H_0 \) cannot be rejected; that is, the two classification variables are independent of each other. Otherwise, reject \( H_0 \) and accept that there is a dependency between the two classification variables.
The correlation coefficient describes the degree of correlation between two variables. This study uses the correlation coefficient \( V \), with the same symbol used in the independence test, and the calculation is shown as the following equation:

\[
V = \sqrt{\frac{X^2}{n \times \min[(R - 1), (C - 1)]}}
\]

(11)

Correlation analysis between the influencing factors and the generation of electronic waybills was conducted. The results are obtained as shown in Table 4.

This analysis shows whether an electronic waybill that is generated has a certain correlation with some influencing factors. The greatest correlation is observed with the declaration and discharge states, duration of stay, and speed. A relatively large correlation is observed with the load states in and out of the electronic fence, carrying, and carriage open. Therefore, the above six influencing variables are chosen as input variables.

4.1.2. Waybill Prediction Result Based on XGBoost. XGBoost is constructed to predict waybill generation. The method is shown in Section 3.1. Six influencing variables are input variables, and whether generation of a waybill is output variable. Several parameters need to be determined in XGBoost. The number of basic learners is the number of decision trees, which is represented by \( n_{\text{estimators}} \); \( max_{\text{depth}} \) is the maximum depth of each tree; \( min_{\text{samples_split}} \) is the minimum number of samples required for internal node subdivision; \( colsample_bytree \) is the percentage of all features used in training each tree. \( scale_{\text{pos_weight}} \) is the weight of positive samples; \( min_{\text{leaf_weight}} \) is the sum of the minimum sample weights of the leaf nodes. \( gamma \) is the penalty coefficient, namely, the minimum loss function decline value required by node partitioning; \( subsample \) is the percentage of subsamples used to train each tree to the total sample; \( reg_{\alpha} \) is the regularization coefficient of \( L_1 \); \( reg_{\lambda} \) is the regularization coefficient of \( L_2 \). In the binary classification task, when the proportion of positive and negative samples is unbalanced, the weight of positive samples is set to achieve a better model effect.

Initialization parameters: \( n_{\text{estimators}} \) equals 100; \( max_{\text{depth}} \) equals 1; \( learning_{\text{rate}} \) equals 0.3; \( subsample \) equals 0.7; \( colsample_bytree \) is set to 0.7; \( min_{\text{leaf_weight}} \) equals 1; \( gamma \) equals 1; \( reg_{\alpha} \) equals 1; and \( reg_{\lambda} \) equals 1. The sample size of the training set is set to 9762, and the sample size of the test set is set to 4184. The model is fitted, and the model accuracy rate is 0.8936. To obtain better results, the parameter tuning process for XGBoost is as follows.

(1) Parameter \( max_{\text{depth}} \) is set from 3 to 10; the step size is set 1; \( min_{\text{child_weight}} \) is set from 1 to 6; and the step size is set 1. The result of parameter tuning is shown in Table 5. The optimal \( max_{\text{depth}} \) is set to 4, and \( min_{\text{child_weight}} \) is set to 1.

(2) Parameter \( gamma \) is set from 0 to 0.5, and the step size is 0.1. The results are shown in Table 6. The best value of \( gamma \) is 0.

(3) The values of the \( subsample \) and \( colsample_bytree \) are set from 0.7 to 1.0. The step size is 0.1. The results are shown in Table 7. The best value of \( subsample \) is 0.9. The best value of \( colsample_bytree \) is 0.9.
Table 4: Results of the correlation analysis.

<table>
<thead>
<tr>
<th>Test</th>
<th>Declaration and discharge states</th>
<th>Duration of stay</th>
<th>Load states in and out of the electronic fence</th>
<th>Carrying</th>
<th>Carriage open</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>5535.128</td>
<td>6640.536</td>
<td>1546.306</td>
<td>2758.526</td>
<td>2575.341</td>
<td>5889.093</td>
</tr>
<tr>
<td>Significance Correlation coefficient $V$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

| $V$ Correlation | | | 0.630 | 0.690 | 0.333 | 0.445 | 0.430 | 0.650 |

Table 5: Results of parameter tuning with variations in $max_depth$ and $min_child_weight$.

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameters</th>
<th>Mean_validation_score</th>
<th>Cv_validation_scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$['max_depth':3,'min_child_weight':1]$</td>
<td>0.90196681</td>
<td>[0.90126997 0.90266393]</td>
</tr>
<tr>
<td>1</td>
<td>$['max_depth':3,'min_child_weight':2]$</td>
<td>0.901761934</td>
<td>[0.90147481 0.9024918]</td>
</tr>
<tr>
<td>2</td>
<td>$['max_depth':3,'min_child_weight':3]$</td>
<td>0.90196681</td>
<td>[0.90166514 0.90286885]</td>
</tr>
<tr>
<td>3</td>
<td>$['max_depth':4,'min_child_weight':1]$</td>
<td>0.904118009</td>
<td>[0.90342448 0.90389344]</td>
</tr>
<tr>
<td>4</td>
<td>$['max_depth':4,'min_child_weight':2]$</td>
<td>0.903608581</td>
<td>[0.90311348 0.90409836]</td>
</tr>
<tr>
<td>5</td>
<td>$['max_depth':4,'min_child_weight':3]$</td>
<td>0.902581438</td>
<td>[0.90208931 0.90307377]</td>
</tr>
<tr>
<td>6</td>
<td>$['max_depth':5,'min_child_weight':1]$</td>
<td>0.903708257</td>
<td>[0.90342448 0.90307377]</td>
</tr>
<tr>
<td>7</td>
<td>$['max_depth':5,'min_child_weight':2]$</td>
<td>0.903196066</td>
<td>[0.90272798 0.90266393]</td>
</tr>
<tr>
<td>8</td>
<td>$['max_depth':5,'min_child_weight':3]$</td>
<td>0.902376562</td>
<td>[0.90229414 0.90249502]</td>
</tr>
<tr>
<td>9</td>
<td>$['max_depth':6,'min_child_weight':1]$</td>
<td>0.90299119</td>
<td>[0.90331831 0.90266393]</td>
</tr>
<tr>
<td>10</td>
<td>$['max_depth':6,'min_child_weight':2]$</td>
<td>0.902888752</td>
<td>[0.90290864 0.90286885]</td>
</tr>
<tr>
<td>11</td>
<td>$['max_depth':6,'min_child_weight':3]$</td>
<td>0.903093628</td>
<td>[0.90270381 0.90348361]</td>
</tr>
</tbody>
</table>

Table 6: Results of parameter tuning with variations in $gamma$.

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameters</th>
<th>Mean_validation_score</th>
<th>Cv_validation_scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$['gamma':0]$</td>
<td>0.904118009</td>
<td>[0.90342448 0.90389344]</td>
</tr>
<tr>
<td>1</td>
<td>$['gamma':0.1]$</td>
<td>0.904118009</td>
<td>[0.90342448 0.90389344]</td>
</tr>
<tr>
<td>2</td>
<td>$['gamma':0.2]$</td>
<td>0.903810695</td>
<td>[0.90393281 0.90368852]</td>
</tr>
<tr>
<td>3</td>
<td>$['gamma':0.3]$</td>
<td>0.903608581</td>
<td>[0.90372798 0.90348361]</td>
</tr>
<tr>
<td>4</td>
<td>$['gamma':0.4]$</td>
<td>0.90350338</td>
<td>[0.90352315 0.90348361]</td>
</tr>
<tr>
<td>5</td>
<td>$['gamma':0.5]$</td>
<td>0.90350338</td>
<td>[0.90352315 0.90348361]</td>
</tr>
</tbody>
</table>

Table 7: Results of parameter tuning with variations in $subsample$ and $colsample_bytree$.

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameters</th>
<th>Mean_validation_score</th>
<th>Cv_validation_scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$['subsample':0.7,'subsample':0.7]$</td>
<td>0.904322885</td>
<td>[0.90311348 0.90553279]</td>
</tr>
<tr>
<td>1</td>
<td>$['subsample':0.7,'subsample':0.8]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>2</td>
<td>$['subsample':0.7,'subsample':0.9]$</td>
<td>0.904322885</td>
<td>[0.90393281 0.90471311]</td>
</tr>
<tr>
<td>3</td>
<td>$['subsample':0.8,'subsample':0.7]$</td>
<td>0.904322885</td>
<td>[0.90311348 0.90553279]</td>
</tr>
<tr>
<td>4</td>
<td>$['subsample':0.8,'subsample':0.8]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>5</td>
<td>$['subsample':0.8,'subsample':0.9]$</td>
<td>0.904322885</td>
<td>[0.90393281 0.90471311]</td>
</tr>
<tr>
<td>6</td>
<td>$['subsample':0.9,'subsample':0.7]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>7</td>
<td>$['subsample':0.9,'subsample':0.8]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>8</td>
<td>$['subsample':0.9,'subsample':0.9]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>9</td>
<td>$['subsample':0.9,'subsample':0.9]$</td>
<td>0.904732637</td>
<td>[0.90432448 0.90389344]</td>
</tr>
<tr>
<td>10</td>
<td>$['subsample':1.0,'subsample':0.7]$</td>
<td>0.904732637</td>
<td>[0.90413675 0.90532787]</td>
</tr>
<tr>
<td>11</td>
<td>$['subsample':1.0,'subsample':0.8]$</td>
<td>0.904732637</td>
<td>[0.90413675 0.90532787]</td>
</tr>
<tr>
<td>12</td>
<td>$['subsample':1.0,'subsample':0.9]$</td>
<td>0.904322885</td>
<td>[0.90393281 0.90532787]</td>
</tr>
<tr>
<td>13</td>
<td>$['subsample':1.0,'subsample':1.0]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
<tr>
<td>14</td>
<td>$['subsample':1.0,'subsample':1.0]$</td>
<td>0.904732637</td>
<td>[0.90432448 0.90389344]</td>
</tr>
<tr>
<td>15</td>
<td>$['subsample':1.0,'subsample':1.0]$</td>
<td>0.903913133</td>
<td>[0.9049898 0.90532787]</td>
</tr>
</tbody>
</table>

(4) Apply regularization to reduce overfitting and adjust the $reg\_alpha$ value. Set $reg\_alpha$ to 0.001, 0.01, 0.1, 1, and 10. The results are shown in Table 8. The best value of $reg\_alpha$ is 0.001.

4.1.3. Model Comparison and Analysis. The optimal results of each model are shown in Table 9. The prediction accuracy of XGBoost is better than that of the decision tree, random forest, and GBDT. Therefore, XGBoost is used to fit the
The XGBoost model is used to predict the test dataset. The confusion matrix is shown in Table 11. The accuracy is 90.057%; thus, the model fitting accuracy is good.

The ROC curve is drawn by using the vertical axis as the true-positive rate (TPR) shown in equation (12) and the horizontal axis as the false-positive rate (FPR) shown in equation (13). The area under the ROC curve is the value of AUC. The closer the AUC value to 1, the better the performance and generalization ability of the model. The ROC curve is obtained as shown in Figure 3. The AUC value is 0.91, which is close to 1, indicating that the prediction model is well constructed.

\[ TPR = \frac{TP}{TP + FN}, \quad (12) \]
\[ FPR = \frac{FP}{TN + FP}. \quad (13) \]

A feature importance analysis can enhance the interpretability of the model, which helps establish model trust and make realistic decisions. The feature importance is obtained based on XGBoost, as shown in Figure 4. The variable \( f_0 \) is the declaration and discharge state; \( f_1 \) is the load state in and out of the electronic fence; \( f_2 \) is acceleration; \( f_3 \) is carrying; \( f_4 \) is speed; \( f_5 \) is carriage open. The declaration and discharge state and load state in and out of the electronic fence variables are the most important. Carriage open and speed are the least important.

4.2. Case Study on Hot Nodes Identification. After the electric waybill is generated, the transport trajectory of working trucks, which is recorded in the waybill, is obtained. With the bottleneck of urban roads, the path node has a complicated traffic environment. To control the important supervision positions on the construction waste transportation path, the time and space characteristics of the transport trajectory are analyzed. The hot nodes of the transport trajectory are developed based on the DBSCAN model. The management of important path nodes in construction waste transportation can not only improve the efficiency of construction waste transportation management but also provide a reference for the selection of construction waste transportation roads.

The thermal map [38] of the operation trajectory of the garbage trucks is shown in Figure 5. For the peak period from 13:00 to 16:00, the spatial distribution of the trajectory of the garbage truck in each period is relatively even. For the trajectory during the period from 20:00 to 21:00, the color of the spatial distribution heatmap is the darkest; therefore, the number of trajectories is obviously greater than that that in other periods. Therefore, it can be concluded that the construction waste transportation task is the heaviest in this period.

**Table 8: Results of parameter tuning with variations in reg_alpha.**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean_validation_score</th>
<th>Cv_validation_scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 {'reg_alpha': 0.001}</td>
<td>0.903913133</td>
<td>[0.90311348 0.90471311]</td>
</tr>
<tr>
<td>1 {'reg_alpha': 0.01}</td>
<td>0.904015571</td>
<td>[0.90311348 0.90491803]</td>
</tr>
<tr>
<td>2 {'reg_alpha': 0.1}</td>
<td>0.904015571</td>
<td>[0.90311348 0.90491803]</td>
</tr>
<tr>
<td>3 {'reg_alpha': 1}</td>
<td>0.904732637</td>
<td>[0.90393281 0.90553279]</td>
</tr>
<tr>
<td>4 {'reg_alpha': 10}</td>
<td>0.896332719</td>
<td>[0.89430561 0.89836066]</td>
</tr>
</tbody>
</table>

**Table 9: Model comparison.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Decision tree</th>
<th>Random forest</th>
<th>GDBT</th>
<th>XGBoost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.8421</td>
<td>0.8769</td>
<td>0.8892</td>
<td>0.9057</td>
</tr>
</tbody>
</table>

**Table 10: Confusion matrix.**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Prediction result</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>TP</td>
</tr>
<tr>
<td>False</td>
<td>FP</td>
</tr>
</tbody>
</table>

**Table 11: Confusion matrix of prediction results.**

<table>
<thead>
<tr>
<th>Prediction result</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waybill generation</td>
<td>2368</td>
<td>333</td>
</tr>
<tr>
<td>No waybill generation</td>
<td>83</td>
<td>1400</td>
</tr>
</tbody>
</table>

**Figure 3: ROC curve.**

\[ TPR = \frac{TP}{TP + FN} \quad (12) \]
\[ FPR = \frac{FP}{TN + FP}. \quad (13) \]
The clustering result of DBSCAN is related to the setting of the neighborhood radius $\epsilon$ and the minimum number of samples $MinPts$. The repeated test method is used to find the best clustering effect. According to experience, fix the minimum number of samples $MinPts$ to 100, and set different neighborhood radii $\epsilon$ to 50 meters, 100 meters, 150 meters, and 200 meters. The clustering results are shown in Table 12. The clustering results are visualized based on a scatter plot [39], as shown in Figure 6. It shows that as the radius of the neighborhood increases, the clustering number increases, and the clustering radius becomes wider. When $\epsilon$ is 50 meters, the clusters are concentrated in two districts of Shenzhen. There are only a few clusters in other districts. When $\epsilon$ is 100 meters, the number of clusters has reached 43 types, with noise accounting for 70.61%. When $\epsilon$ is 150 meters and 200 meters, there are more clusters, the proportion of noise is less than 65%, and the clustering results are broad. The number of clusters is low, which is not in line with delicacy management. Therefore, when $\epsilon$ is selected as 100 meters, the result is more reasonable.

Then, $\epsilon$ is fixed to 100 meters, and the minimum number of samples $MinPts$ is set to 50, 100, 150, and 200. The clustering results are shown in Table 13 and Figure 7. It shows that as the parameter $MinPts$ increases, the number of clusters decreases. When the minimum number of samples $MinPts$ is 50, there are too many clusters, reaching 109. When $MinPts$ is greater than 100, the number of clusters is less than 30, which means that the noise ratio is large. Therefore, considering the minimum number of samples $MinPts$ and the neighborhood radius $\epsilon$, the parameter values selected in this study are $MinPts$ equal to 100 and $\epsilon$ equal to 100 meters.

<table>
<thead>
<tr>
<th>($\epsilon$, $MinPts$)</th>
<th>Clustering number</th>
<th>Ratio of noise points</th>
</tr>
</thead>
<tbody>
<tr>
<td>(50, 100)</td>
<td>30</td>
<td>0.8057</td>
</tr>
<tr>
<td>(100, 100)</td>
<td>43</td>
<td>0.7061</td>
</tr>
<tr>
<td>(150, 100)</td>
<td>51</td>
<td>0.6365</td>
</tr>
<tr>
<td>(200, 100)</td>
<td>60</td>
<td>0.5336</td>
</tr>
</tbody>
</table>
Table 13: Clustering results with a minimum number of samples of 100.

<table>
<thead>
<tr>
<th>($\epsilon$, $\text{MinPts}$)</th>
<th>Clustering number</th>
<th>Ratio of noise points</th>
</tr>
</thead>
<tbody>
<tr>
<td>(100, 50)</td>
<td>109</td>
<td>0.5196</td>
</tr>
<tr>
<td>(100, 100)</td>
<td>43</td>
<td>0.7061</td>
</tr>
<tr>
<td>(100, 150)</td>
<td>25</td>
<td>0.7964</td>
</tr>
<tr>
<td>(100, 200)</td>
<td>12</td>
<td>0.8695</td>
</tr>
</tbody>
</table>
Due to the density-based clustering analysis, the shape of the clustering results is often irregular. To better represent the location of the hot node, the center position of the cluster is used to represent the location of the hot node. The calculation for the construction waste transportation hot node is shown in the following equation:

\[
(x', y') = \left( \frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right),
\]

where \((x_i, y_i)\) is the position of the \(i\)-th garbage truck, \(n\) is the data number of the cluster, and \((x', y')\) is the central location of the cluster, which is the hot node.

According to the clustering results, calculate the latitude and longitude of the center point of each cluster and mark it on the map, as shown in Figure 8. For hot nodes, managers should carry out key management to prevent irregularities in garbage trucks to reduce the impact of construction waste transportation on urban traffic.

5. Conclusion

A large amount of construction waste is generated in urban construction. Electronic waybill technology is used to manage the process of the generation, transportation, and disposal of construction waste to improve the efficiency of management. Transport trajectories, as important control roads in traffic management, occupy an important position in construction waste transportation management. To this end, the transport trajectory of working trucks is analyzed in this study. Working trucks are obtained from electronic waybills. Improving the prediction accuracy of electronic waybill generation can improve the identification accuracy of working trucks. Therefore, the generation of electronic waybills is predicted. After obtaining the transport trajectory of working trucks, the spatial-temporal characteristics of the transport trajectory are analyzed and hot nodes are identified. Research on trajectory analyses can improve the quality of construction waste transportation management.

A case study of Shenzhen is introduced. First, a correlation analysis on the influencing factors of electronic waybill generation is conducted, and six influencing factors are found. To predict the generation of waybill, the decision tree, random forest, GBDT, and XGBoost methods are used. According to the model comparison, XGBoost is better at fitting the prediction model, and the accuracy can reach 90.06%. Then, the trajectory data during the peak period of construction waste transportation are clustered based on the DBSCAN model. The results show that the important path nodes can be found and visualized. It is helpful to identify important control positions on the transportation path of construction waste, which will improve the efficiency of transportation management. The method proposed in this study can meet the requirements of engineering practice.

Certain shortcomings were observed in this study. Hot nodes are found based on the trajectory of garbage trucks. However, path nodes that have high traffic accidents due to narrow roads and crowded people may not be considered. In future studies, road environmental factors will be added to the node analysis. Moreover, the other prediction models are tested to improve the accuracy of waybill generation. Electronic waybill technology will be improved to manage the transportation process automatically.

Data Availability

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
Acknowledgments

This article was sponsored by National Key R&D Program of China (2018YFC0706005).

References


[31] M. Chikaraishi, P. Garg, V. Varghese et al., "On the Possibility of Short-Term Traffic Prediction during Disaster with


