Article

Expressway Speed Prediction Based on Electronic Toll Collection Data

Fumin Zou 1, Qiang Ren 1,*, Junshan Tian 1, Feng Guo 2, Shiben Huang 1, Lyuchao Liao 3 and Jinshan Wu 1

1 Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China; fmzou@fjut.edu.cn (F.Z.); 2191901003@qq.com (J.T.); 2201901025@qq.com (S.H.); 2191905037@qq.com (J.W.)
2 College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China; n180310004@fzu.edu.cn
3 Fujian Provincial Big Data Research Institute of Intelligent Transportation, Fujian University of Technology, Fuzhou 350118, China; achao@fjut.edu.cn
* Correspondence: rq809887266@gmail.com; Tel.: +86-133-6843-4080

Abstract: Expressway section speed can visually reflect the section operation condition, and accurate short time section speed prediction has a wide range of applications in path planning and traffic guidance. However, existing expressway speed prediction data have defects, such as sparse density and incomplete object challenges. Thus, this paper proposes a framework for a combined expressway traffic speed prediction model based on wavelet transform and spatial-temporal graph convolutional network (WSTGCN) of the Electronic Toll Collection (ETC) gantry transaction data. First, the framework pre-processes the ETC gantry transaction data to construct the section speeds. Then wavelet decomposition and single-branch reconstruction are performed on the section speed sequences, and the spatial features are captured by graph convolutional network (GCN) for each reconstructed single-branch sequence, and the temporal features are extracted by connecting the gated recurrent unit (GRU). The experiments use the ETC gantry transaction data of the expressway from Quanzhou to Xiamen. The results indicate that the WSTGCN model makes notable improvements compared to the model of the baseline for different prediction ranges.

Keywords: expressway speed prediction; the Electronic Toll Collection; graph convolutional network; gated recurrent unit; wavelet transform

1. Introduction

With its high capacity and low time cost, the expressway has become the preferred way to travel between cities [1]. Due to social and economic development, traditional traffic management techniques struggle to cope with the increasing traffic pressure, and there is an urgent need to develop Intelligent Transportation Systems (ITS) for expressways. With the accurate prediction of traffic information from ITS, travelers can develop reasonable travel routes before departure and improve the efficiency of travel, and road management departments can effectively conduct traffic guidance and alleviate traffic congestion and other problems based on reliable road traffic information [2]. In recent years, China’s expressway ETC system has realized the networking of 29 provinces nationwide, built a total of 24,588 sets of ETC gantry systems, renovated 48,211 ETC lanes, and averaged nearly one billion ETC gantry transaction data per day [3], which has further improved the efficiency of expressways. The transaction data collected by the ETC gantry system can record the travel information of almost every vehicle on the expressway, and compared with detector data [4,5] and floating car data [6,7], the ETC gantry transaction data are more comprehensive and reliable, covering the expressway road network. Therefore, to further improve the service quality of the expressway ETC system, it is of great theoretical significance and practical value to research the traffic speed prediction based on ETC gantry transaction data [8].

Electronics 2022, 11, 1613. https://doi.org/10.3390/electronics11101613
https://www.mdpi.com/journal/electronics
Several domestic and foreign researchers have invested in the field of traffic prediction research [9]. The three main types of traffic prediction methods are statistical learning models, shallow machine learning and deep learning. Statistical learning models. For example, Emami et al. [10] proposed a Kalman filter-based method for traffic flow prediction. Xu et al. [11] proposed an Autoregressive Integrated Moving Average (ARIMA) and Kalman filter method for predicting road traffic status. Statistical learning models are usually performed under the assumption of linearity, which does not reflect the nonlinear feature of traffic speed well, and also cannot handle high-dimensional data well due to its high complexity. Shallow machine learning can solve the above problems well, for example, Hu et al. [12] proposed a Support Vector Regression (SVR)-based method for estimating the average speed of expressway sections to overcome the sparse density of existing expressway vehicle detectors. Evans et al. [13] performed road section state prediction based on Random Forest (RF), and with different prediction ranges and training data amounts, the algorithm achieved the best results compared with others. Sun et al. [14] used a method to dynamically adjust the K-nearest neighbor (KNN) parameters for traffic flow prediction and achieved good results in different periods. However, the performance of shallow machine learning methods depends heavily on the artificially designed features and they usually fail to produce the best results for prediction tasks with complex regularities and complicated factors. In recent years, deep learning models have shown their superior predictive capabilities. Deep learning models can automatically extract features and capture the correlation of data, so a large number of deep learning models are used for traffic prediction. Since traffic data can be represented as time series data, GRU [15] and LSTM [16] are gradually used for speed prediction. Fu et al. [17] used wavelet transform to decompose the original data and then constructed GRU and Autoregressive Moving Average (ARMA) models to predict low-frequency and high-frequency sequences. Despite the impressive capability of these methods in temporal modeling, accurate traffic prediction is still limited because of the lack of consideration of the spatial characterization of traffic data. To solve this problem, Lu et al. [18] proposed a spatial-temporal deep learning network (ST-TrafficNet) for traffic flow forecasting, which is capable of capturing high-dimensional temporal features while also extracting latent spatial features. Bogaerts et al. [19] used a combination of CNN and LSTM to construct a spatio-temporal recurrent convolutional neural network to effectively extract temporal features and spatial feature variations of traffic speed and achieve high accuracy traffic speed prediction. However, CNN is designed for Euclidean spatial structure, and for the actual expressway road network structure, CNN is unable to capture the spatial correlation completely. Zhao et al. [20] proposed a temporal graph convolutional network for traffic prediction, in which the combination of GCN and GRU goes to capture the spatial and temporal features of traffic flow. Pan et al. [21] proposed the dual-channel based graph convolutional network (DC-STGCN) model to fully extract the spatio-temporal characteristics between traffic flows, and achieved good results in long-term prediction.

Regarding expressway speed prediction, most of the data used in this paper are vehicle detector data and floating car data. However, there are shortcomings, such as low density of vehicle detectors, high damage rate, and incomplete objects for floating car data. To solve these issues, this paper considers the nearly full sample of ETC gantry transaction data and use the WSTGCN model to predict expressway speeds, which can effectively eliminate the section speeds and extract the spatio-temporal characteristics between ETC gantries. The main contributions of this work as follows.

- A data pre-processing method on ETC gantry transaction data is designed. The fusion of expressway network topology data, ETC and manual toll collection (MTC) transaction data constitutes spatio-temporal origin-destination (OD) data. Anomaly cleaning, missing repair and vehicle travel time statistics are performed on OD data, and a vehicle travel time outlier detection algorithm is proposed to eliminate outlier samples. In this way, the speed of the expressway section is constructed.
The proposed WSTGCN model consists of wavelet transform, GCN and GRU. It reduces the disturbance of section speed and also captures the spatio-temporal correlation section speed.

The proposed WSTGCN model is evaluated on the ETC gantry transaction data of the Quanzhou-Xiamen Expressway. The results show that the model has the best prediction of section speed compared with the baseline method. Furthermore, the accuracy is still higher than that of the baseline prediction model in different prediction ranges.

The rest of the paper is organized as follows. The concepts related to expressways are introduced and problem description in detail in the “Preliminary” section. The model construction for expressway speed prediction is described in detail in the “Methodology” section. In the “Experimental Results and Analysis” section, the WSTGCN model is evaluated using ETC gantry transaction data from Fujian Province, and finally we present the ‘Conclusions” of our paper.

2. Preliminary

2.1. Related Concepts

Definition 1. Each ETC gantry of the expressway is called a Node, and two adjacent Nodes on the road compose an expressway section, which is referred to as \( QD = \{ Q, \text{Distance} \} \), \( Q = \langle \text{Node}_1, \text{Node}_2 \rangle \), where \( \text{Node}_1 \) is the start of the section, \( \text{Node}_2 \) is the end of the section, and Distance is the actual distance of the section.

Definition 2. Expressway road network, all QD within the research area of expressway form expressway road network, referred to as \( LW = \{ QD_1, \ldots, QD_n \} \).

Definition 3. Vehicle Trajectory, the sequence of nodes arranged in chronological order formed by a vehicle on the ETC gantry on the expressway is called \( \text{Traj} = \{ \text{Node}_1, \ldots, \text{Node}_n \} \), where \( \text{Node}_1 \) is called the trajectory start point, and \( \text{Node}_n \) is called the trajectory end point.

Definition 4. The average speed of a vehicle passing through a section is called the average vehicle speed. The calculation method is as follows:

\[
v_i = \frac{\text{Distance}}{t_2 - t_1}
\]

where \( t_1 \) represents the time when the vehicle passes through the starting point of the section, and \( t_2 \) represents the time when the vehicle passes through the end point of the section.

Definition 5. The average speed of vehicles passing through the same section in a certain period of time is called section speed, and the calculation method is shown in Equation (2)

\[
V = \frac{\sum_{i=1}^{n} v_i}{n}
\]

where \( v_i \) represents the average vehicle speed of the ith vehicle, \( i \) is the ith vehicle passing through a certain section within a certain period of time, and \( n \) is the nth vehicle passing through a certain section within a certain period of time.

Definition 6. The time difference between a vehicle passing through a certain section is called the vehicle travel time, and the calculation method is shown in Equation (3)

\[
\Delta t_i = t_2 - t_1
\]

where \( t_2 \) represents the time of passing a gantry after a certain section, and \( t_1 \) represents the time of passing a gantry before a certain section.
2.2. Problem Description

The expressway road network can be abstracted as a graph. Generally, the unweighted graph \( G = (\varphi, E) \) can be used to represent the topology of the expressway road network, where \( \varphi \) represents the set of all nodes on the expressway road network, \( \varphi = \{ \varphi_1, \varphi_2, \cdots, \varphi_N \} \), and \( N \) represents the number of Nodes. \( E \) represents the set of interconnected edges between Nodes, all the connection information between Nodes is in the adjacency matrix \( A \in \mathbb{R}^{(N-1) \times (N-1)} \), and there are only 0 and 1 elements in the adjacency matrix, where 0 indicates that there is no connection between Nodes, and 1 indicates that there is a connection between Nodes. The section speed can be regarded as the attribute feature of the expressway road network Node, which is represented by the feature matrix \( X \in \mathbb{R}^{(N-1) \times P} \), where \( P \) is the number of attribute features of the Node, the length of the historical time series. \( X_t \in \mathbb{R}^{(N-1) \times i} \) represents the section speed of all sections when the time section is \( i \).

Therefore, the problem of expressway speed prediction is to learn a mapping function \( F \) based on the feature matrix \( X \) of the section speed in the past under the topology graph \( G \) of the expressway road network to predict the section speed at \( T \) times in the future.

\[
X_{t+T} = F((X_{t-n}, \cdots, X_{t-1}, X_t); G) \tag{4}
\]

where \( t \) is the time interval, \( n \) is the length of the historical time series, and \( T \) is the length of the time series to be forecasted.

3. Methodology

3.1. Overview of the Overall Framework

This paper is based on the WSTGCN model to predict expressway section speed, which is mainly divided into four modules: expressway road network spatio-temporal OD data construction module, OD data pre-processing module, spatio-temporal feature extraction module, and output module. Figure 1 shows the whole framework structure of expressway speed prediction. The expressway road network spatio-temporal OD data construction module consists of expressway road network topology data, ETC transaction data and MTC transaction data. According to the expressway road network topology data, ETC, MTC transaction data in the OBU Plate and Flag ID group iterations, and then Trade time will be sorted to form the spatio-temporal OD data set. The data pre-processing module includes a data interpolation module, vehicle travel time construction module, travel time abnormality detection module, and section speed generation module. After interpolating the missing data of some trajectories, the vehicle travel time of each section is constructed, and then the anomaly is detected using the vehicle travel time outlier detection algorithm, finally, the section speed data set is constructed. The spatio-temporal feature extraction module includes wavelet transform, GCN, and GRU. The multi-scale wavelet decomposition is applied to the section speed time series data, decomposed and reconstructed to obtain the section speed after single branch reconstruction, and then GCN is used to capture the spatial feature information of section speed, and GRU is used to capture the temporal feature information of section speed. The final output module, which outputs the numerical summation of the predicted values of each reconstructed single-branch series, obtains the overall speed prediction results considering the spatio-temporal characteristics.
3.2. Data Pre-Processing

Vehicles enter the expressway through the ETC channel and MTC channel of the expressway toll station, and the expressway ETC gantry system can record the driving information of vehicles entering from the ETC channel and MTC channel at the same time. Therefore, ETC gantry transaction data are more complete. ETC gantry transaction data includes ETC transaction data and MTC transaction data. According to the ETC gantry transaction data statistics of this experiment, the percentages of ETC transaction data and MTC transaction data are shown in Figure 2.
3.2.1. Raw Data Cleaning

In the process of ETC gantry transaction data collection, the following three main abnormal problems exist in the collected ETC gantry transaction data, due to the influence of factors beyond control such as equipment abnormalities, wireless crosstalk, and bad weather, as shown in Figure 3. (1) Data redundancy. Duplication between multiple sets of data. (2) Missing data. The problem of data not being collected effectively occurs. For example, fields such as date, time, and vehicle type are missing at the entrance and exit station. (3) Data errors. Data records that do not match the normal traffic rules, such as the date of the entrance station being later than the date of the exit station, and the wrong entrance and exit numbers, which cannot correspond to the actual toll stations. These abnormal data greatly reduce the value of ETC big data-mining applications. To reduce the impact of erroneous data on the accuracy of the established prediction model and increase the reliability of prediction, such data will be removed.

3.2.2. Vehicle Travel Time Construction

After the abnormal data of ETC gantry transaction data are eliminated, the travel trajectory of each vehicle is constructed by the time sequence. Using the ETC gantry topology data of the expressway road network, the ETC gantry search is performed for each vehicle’s travel trajectory, traversing two adjacent ETC gantries in the vehicle travel trajectory, checking the two adjacent gantry topology relationships and whether they exist in the ETC gantry topology data of the expressway road network. If it exists, the travel time of the vehicle through the section is calculated directly. If it does not exist, road section is searched with these two gantries, the driving trajectory of the vehicle is interpolated, the average speed of the vehicle through the road section according to the search result is calculated, and this average speed is taken as the average speed of all the sections between through these two gantries, through the distance of the two adjacent gantries, the travel time of this section can be derived. The specific construction method is shown in Algorithm 1.
Figure 3. Types of abnormal data (** indicates that the data was desensitized) (a) Data redundancy (b) Missing data (c) Data errors.

Algorithm 1 Vehicle travel time construction algorithm.

Input:
Vehicle trajectory data Traj; Expressway road network topology data LW;

Output:
Output Vehicle travel time data $\Delta t$;

1. $\text{Traj} = \{\text{Node}_1, \cdots, \text{Node}_n\}$, $\text{LW} = \{QD_1, \cdots, QD_n\}$, $QD = \{Q, \text{Distance}\}$;
2. for $i = 0$ to $i = n - 1$ do;
3. $\text{Node}_i, \text{Time}_i, \text{Node}_{i+1}, \text{Time}_{i+1}$ //extract information of adjacent nodes;
4. $\Delta t_i = \text{Time}_{i+1} - \text{Time}_i$ //compute the time difference of adjacent nodes;
5. $Q = (\text{Node}_i, \text{Time}_i, \text{Node}_{i+1}, \text{Time}_{i+1})$ //save the information of adjacent nodes;
6. $\Delta t = (Q, \Delta t_i)$ //save the vehicle passage time data;
7. If $\text{Node}_i$ and $\text{Node}_{i+1}$ in LW //if adjacent nodes are in topological data;
8. $\Delta t = \Delta t(Q, \Delta t_i)$ //the vehicle passage time data remains unchanged;
9. Else $\text{Node}_i$ and $\text{Node}_{i+1}$ not in LW;
10. $\text{distance} = \{\}$;
11. $\{\text{Node}_i, \cdots, \text{Node}_n\} \leftarrow$ shortest path LW //search for the shortest path;
12. $\{\text{distance}_i, \cdots, \text{distance}_n\} \leftarrow$ LW //the shortest distance between gantries to distance;
13. $\nu = \text{Distance} \div \nu$ //calculate the speed of the front and back gantry;
14. $\{\text{Node}_i, \cdots, \text{Node}_n\} = \nu$ //add speed attributes to adjacent gantries;
15. $\Delta t_i = \text{distance} \div \nu$ //calculate the time difference of adjacent nodes;
16. $\text{Node}_j, \text{Node}_{j+1}, \text{Time}_j, \text{Time}_{j+1}, \Delta t_j$ //extract the information of adjacent nodes in the shortest path
17. $\Delta t = (Q, \Delta t_j)$ //replace the original pass time and generate a new pass time;
18. return $\Delta t$;
3.2.3. Vehicle Travel Time Outlier Detection Algorithm

After constructing the vehicle travel time data, the data that objectively exist are reasonable. However, they contain some of the ETC gantry transaction data of abnormal driving behavior, for example, if a vehicle’s travel time is too long or too short compared with the normal situation of similar models. Therefore, a vehicle travel time outlier detection algorithm is constructed to further reject such data. This algorithm is a combination of the outlier information detection algorithm in the literature [22] and the outlier elimination algorithm in the literature [23]. In the expressway ETC gantry transaction data, there are only a few very few cases where the vehicle travel time is shorter than the normal value, and most of the outliers are long vehicle travel times, resulting in an asymmetric error interval. If only the outlier information detection algorithm is used, the 75% quantile value is relatively high, while the 25% quantile value is closer to the sample mean, which will not be able to eliminate the data where the vehicle travel time is much lower than the sample mean. Therefore, combining the two methods can solve this problem well. The basic idea of the vehicle travel time outlier detection algorithm is to use both upper and lower limits of the box line diagram and the centroid threshold of the statistical distribution of distance data for outlier detection, to determine the threshold interval for abnormal travel time data filtering, and to eliminate the data outside this threshold, and then to quickly filter out abnormal data in the massive ETC gantry transaction data, as shown in Figure 4.

$$t_{down} = \max(t_{25\%} - 1.5 \times (t_{75\%} - t_{25\%}), t_{mean} - 2\sigma)$$  \hspace{1cm} (5)

$$t_{up} = \min(t_{75\%} + 1.5 \times (t_{75\%} - t_{25\%}), t_{mean} + 2\sigma)$$  \hspace{1cm} (6)

$$t_{25\%}$$ means the time greater than 25% of the vehicle travel time, $$t_{75\%}$$ means the time greater than 75% of the vehicle travel time, $$t_{mean} = \frac{1}{N}\sum_{i=1}^{N} t_i$$ is the mean value of vehicle travel time, and $$\sigma$$ is the standard deviation of vehicle travel time. The final vehicle travel time is valid interval $$\Delta t \in [t_{down}, t_{up}]$$. If a vehicle passes through a section in a certain time period with a vehicle travel time within $$\Delta t$$, the average vehicle speed of the vehicle passing through the section is directly generated, and the section speed is generated with a statistical window of 15 min.

![Figure 4](image)

**Figure 4.** Principle diagram of travel time outlier detection (a) Upper and lower limits are outside the normal distribution (b) Upper and lower limits are inside the normal distribution (c) Lower limit is outside the normal distribution and upper limit is inside the normal distribution (d) Lower limit is inside the normal distribution and upper limit is outside the normal distribution.

3.3. Spatio-Temporal Feature Extraction

Based on the GCN-GRU model, wavelet transform is used to capture the spatio-temporal trend of expressway traffic speed by decomposing and reconstructing the ex-
pressway traffic speed. The structure of the prediction model is shown in Figure 5, which contains three parts: (a) wavelet transform (b) GCN (c) GRU.

**Figure 5.** WSTGCN model.

### 3.3.1. Wavelet Transform

The expressway ETC gantry transaction data generates a lot of noise due to its periodic volatility, and data containing noise is fatal for speed prediction. In real traffic analysis, it is known that real speed signals are usually low-frequency speed signals or relatively stable speed signals, while noisy signals are more high-frequency speed signals [24]. Therefore, with the help of the theory related to wavelet transform, the calculated section speed signals are filtered out of the noise signals to obtain relatively accurate section speed data. To separate the low-frequency part and the high-frequency part of the original signal, Mallet et al. proposed a multiscale decomposition and reconstruction algorithm for the repair signal, the principle of which is shown in Figure 6.

**Figure 6.** Wavelet transform principle.

\[
A_j[f(t)] = \sum_k H(2t - k)A_{j-1}[f(t)] \\
D_j[f(t)] = \sum_k G(2t - k)A_{j-1}[f(t)]
\]
In the formula, \( t \) is the time series number of the time series data, \( t = 1, 2, \cdots n \), \( f(t) \) is the original signal, \( j \) is the number of layers of decomposition. \( H, G \) are wavelet decomposition filters in the time domain, \( A_j \) is the wavelet coefficient of the low-frequency part of the signal \( f(t) \) in the \( j \)th layer, and \( D_j \) is the wavelet coefficient of the high-frequency part of the signal \( f(t) \) in the \( j \)th layer. The decomposed signal can be reconstructed using Equation (9).

\[
f(t) = D_1 + \cdots + D_j + A_j
\]  

(9)

In the expressway speed prediction, the original section speed data consists of a set of non-smooth time series data. In many wavelet transform functions, the sym wavelet function is a linear phase, approximately symmetric and double orthogonal function. The smoothness is better, the calculation is simpler, and it has achieved good results in related research [25,26]. The sym5 is one of the commonly used wavelets in the sym wavelet group. Therefore, in this paper, the sym5 wavelet is chosen as the basis function. The number of decomposition layers cannot be too large or too small. If the number of decomposition layers is too large, it will reduce the variation pattern and trend of the section speed series. If the number of decomposition layers is too small, the signals with different frequency characteristics in the original section speed signal cannot be separated effectively. According to the existing research on wavelet transform for noise reduction of time series data [27], the number of decomposition layers is set to 3. The high-frequency part of the decomposed signal in each layer is processed using a threshold function. Finally, the low-frequency speed signal of the last layer after decomposition is reconstructed with the high-frequency speed signal after the threshold in each layer to obtain the noise-reduced section speed data. The decomposition results are shown in Figure 7, and the specific method is described in Algorithm 2.
where $g$ polynomial form of \als are used to approximate the calculation. This paper adopts the simplified first-order complexity of $g$ mapping $g$ can be converted into the frequency domain inner product form as:

$$
\text{Algorithm 2: Wavelet transform algorithm.}
$$

**Input:**
the set of section speed time series $v$;

**Output:**
the set of section speed time series $v'$;

1: $v = \{v_1, \cdots, v_n\}, ca = [\cdots], cd = [\cdots], reca = [\cdots], recd = [\cdots]$;
2: Select the sym5 wavelet as the basis function;
3: $j = 3$ //the number of decomposition layers is specified as 3 layers;
4: for $i$ to range($j$);
5: $ca \leftarrow ca_i + \cdots + ca_j$ //store the trend signal;
6: $cd \leftarrow cd_i + \cdots + cd_j$ //store the noise signal;
7: For $i$ to $ca$;
8: $reca \leftarrow ca_1 + \cdots + ca_i$ //store the reconstructed trend signal;
9: For $i$ to $cd$;
10: $recd \leftarrow cd_1 + \cdots + cd_i$ //store the reconstructed noise signal;
11: $v' \leftarrow reca$ //wrap the reconstructed trend signal into $v'$;
12: **return** $v'$;

3.3.2. Graph Convolutional Networks (GCN)

The ETC gantry of the expressway has different topological relationships in different sections, and the mutual influence between the ETC gantry with different topological relationships must be different. If the topological relationship between the ETC gantry can be fully extracted and used, the speed prediction will be more accurate. Ordinary CNN can only handle Euclidean spatial data with regular structure, and cannot handle irregular non-Euclidean spatial data. Therefore, the literature [28] proposed the GCN model to deal with non-Euclidean spatial data very well. The spatial distribution of the ETC gantry of the expressway is a non-Euclidean spatial structure, so the GCN model is used to model the spatial distribution of the ETC gantry of the expressway. We treat the data as signals on a spectrogram, and process the signal on the graph to capture meaningful patterns and features in space. The connection relationship and mutual influence of the graph are represented by the Laplacian matrix of the graph. The Laplace matrix of a graph is defined as:

$$
L = D - A
$$

The regularized Laplacian matrix is:

$$
L = I_n - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}
$$

where $I_n \in R_{N \times N}$ is the identity matrix, and the degree matrix $D_N = \sum A_{ij}$. Decompose $L$ into eigenvalues to obtain $L = U\Lambda U^T$, $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix composed of eigenvalues of $L$, $U = \{u_1, \cdots, u_N\}$ is an orthonormal matrix consisting of the standard orthonormal eigenvectors of $L$. For a signal input $X \in R_N$, the Fourier transform in the figure is $\hat{x} = U^T x$, and its inverse Fourier transform is $x = U^T \hat{x}$. The convolution operation of the convolution kernel $g$ and the input signal $x$ in the time domain can be converted into the frequency domain inner product form as:

$$
g \ast x = U((U^T g) \odot (U^T) x) = U_{g0}(A)U^T x
$$

where $g_0(\Lambda) = U^T g = \text{diag}(\Theta)$, $\odot$ represents the Hadamard product, and $U^T g$ means mapping $g$ to the frequency domain space based on $U$. Due to the high computational complexity of $g_0$, the hierarchical linear model constraints [29] and Chebyshev polynomials [30] are used to approximate the calculation. This paper adopts the simplified first-order polynomial form of $g \ast x$:

$$
g \ast x = U_{g0}U^T x \approx \Theta(I_n + D^{-\frac{1}{2}}AD^{-\frac{1}{2}})x
$$
There exists $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} = I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, where $\tilde{A} = I_n + A$ and $\tilde{D} = \sum_i \tilde{A}_{ii}$, so the output of layer $l$ is:

$$X^l = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X^{l-1} W^{l-1})$$ (14)

$A$ represent an adjacency matrix, which is used to represent the connection relationship between expressway nodes. Each row in $A$ represents a section, and each value in $A$ represents the connection between sections. $\tilde{A} = A + I$ in the matrix is to prevent the ETC gantry of the expressway from being unable to transmit its characteristic information when capturing the characteristic information (section speed) of the adjacent nodes. $D$ is the degree matrix, $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ is to prevent the gradient from exploding or disappearing when the gantry propagates feature information layer by layer, which makes it impossible to perform the next step of training. $W^{l-1}$ is the weight in the GCN model, $X^{l-1}$ represents the feature matrix of the section speed, each row represents a different section, each column represents the section speed of the same time interval, $\sigma$ represents an activation function.

As shown in Figure 8, assuming that each node in the figure represents the expressway ETC gantry, the essence of the GCN model is actually to capture the linear combination of the adjacent node features of the gantry and its own single node features. Therefore, Node1 can obtain the spatial characteristics of itself and surrounding nodes through the GCN model.

Figure 8. Topological relationship between gantry to obtain spatial feature expressway.

3.3.3. Gated Recurrent Unit (GRU)

Currently, RNN models in neural network models are commonly used to process sequence data. However, the traditional RNN has the disadvantages of gradient disappearance and gradient explosion in the training process [31]. To solve this problem, GRU and LSTM were proposed as variants of RNNs. However, compared with LSTM, the GRU model has the advantages of simple structure, fewer parameters, and a short training time. Therefore, the GRU model was selected to obtain temporal features from the section speed, and its structure is shown in Figure 9.

Figure 9. GRU basic structure.
There are two gate units in the hidden layer of GRU, the reset gate \( r_t \) and the update gate \( z_t \). \( z_t \) indicates how much the state information of the previous moment is transferred to the current state, and \( r_t \) indicates how much the state information of the previous moment is ignored. The calculation process of GRU is as follows:

\[
\begin{align*}
\hat{H}_t &= \sigma(W_h \cdot [r_t \times H_{t-1}, x_t]) \\
H_t &= (1 - z_t) \times h_{t-1} + z_t \times \hat{H}_t
\end{align*}
\]

Equations (15) and (16) show how to set the update gate \( z_t \) and reset gate \( r_t \). \( W_z \) represents the weight of the reset gate \( r_t \), \( \sigma \) represents the activation function, \( x_t \) represents the section speed at the current moment, and \( H_{t-1} \) represents the hidden state at the time \( t-1 \).

Equation (17) indicates that the output of the reset gate at the current time is multiplied by the hidden state of the previous time, and then the candidate hidden state is calculated through the full connection layer of the activation function. Equation (18) represents the update gate \( z_t \) that calculates the current time, the hidden state \( H_{t-1} \) at the previous time point, and the weighted average of the candidate hidden states \( \hat{H}_t \) at the current time to calculate the most probable state \( H_t \).

In general, this paper uses the historical \( n \) time series data of section speed to obtain the section speed of \( t \) time by using the hidden state of \( t-1 \) time and the current section speed as input through the GRU model. The model captures the section speed at the current moment, while still maintaining the changing trend of historical traffic information, and obtains the dynamic time change characteristics of the section speed.

4. Experimental Results and Analysis

4.1. Data Description and Pre-Processing

The experimental data are mainly divided into two types of data, one is the ETC gantry transaction data of Fuzhou South to Xiamen North Expressway in Fujian Province for 30 days from 1 to 30 June 2020 from Fujian Expressway Information Technology Co., Ltd. (2F, Building 1, No.27 Jinji Shan Road, Jinan District, Fuzhou City, China). which mainly contains transaction 103 dimensions of data, such as transaction identifier, trade time, gantry number, OBU plate, OBU status and user type, with a total of about 20.53 million data samples. The main attributes of the ETC gantry transaction data used in this paper are shown in Table 1. Second, according to the longitude and latitude coordinates of the ETC gantry, the section distance is crawled by Amap, and the topological relationship data of the expressway ETC gantry is generated, which includes the name of the ETC gantry in different sections and the actual section distance.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Examples</th>
<th>Attribute Name</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade ID</td>
<td>340,119...2698</td>
<td>OBU Plate</td>
<td>Blue Fujian A12345</td>
</tr>
<tr>
<td>Trade Time</td>
<td>2020/6/1 21:35:51</td>
<td>Vehicle Class</td>
<td>1</td>
</tr>
<tr>
<td>Flag ID</td>
<td>352,305</td>
<td>Enter Time</td>
<td>2020/6/1 21:27:35</td>
</tr>
<tr>
<td>Flag Type</td>
<td>0</td>
<td>Enter Station</td>
<td>2507</td>
</tr>
<tr>
<td>Flag Index</td>
<td>1</td>
<td>OBU ID</td>
<td>12A2B3F8</td>
</tr>
</tbody>
</table>

Match the ETC transaction data with the topological data to construct the vehicle travel time. Due to the existence of factors such as vehicles entering the service area or vehicles breaking down, a certain amount of abnormal data will be generated. By taking 15 min as the statistical window for each section, the vehicle travel time outlier detection algorithm is used to detect outliers, eliminate abnormal data and retain correct data. The
average speed of each vehicle passing through each section is calculated, and then the speed dataset of the expressway section is constructed with a 15-min interval, and the main attributes are shown in Table 2.

Table 2. Expressway section speed data attribute table.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Flag ID</td>
<td>354,505</td>
</tr>
<tr>
<td>After Flag ID</td>
<td>354,507</td>
</tr>
<tr>
<td>Speed (km/h)</td>
<td>98.46</td>
</tr>
<tr>
<td>Time Window</td>
<td>2020-06-01 13:15:00</td>
</tr>
</tbody>
</table>

In this paper, min-max normalization is used to map the data to the [0, 1] interval, and the normalization formula is shown in Equation (19):

\[
z = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]  

(19)

In Equation (19), \(z\) represents the normalized data, \(x\) represents the original data, \(x_{\text{min}}\) is the minimum value of \(x\), and \(x_{\text{max}}\) is the maximum value of \(x\). According to the above data pre-processing, the section speeds of 16 sections travel in both directions every 15 min from 00:00 to 24:00. Then, a time series of section speed was generated, with 96 data samples per day and a total of 2880 data samples in 30 days. The first 80% of the 30 days of section speed data were used as the training set and the remaining 20% as the test set. The section speeds were predicted for the next 15 min, 30 min, and 45 min.

4.2. Evaluation Indicators

A total of 5 metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Accuracy, Coefficient of Determination (\(R^2\)) and Explained Variance Score (\(\text{Var}\)), were used in the experiments, and they were used to compare and evaluate the prediction results of the models. Where \(y_i\) is the actual section speed, \(\hat{y}\) is the predicted section speed, \(\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i\), and \(N\) is the sample size.

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2}
\]

(20)

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}|
\]

(21)

\[
\text{Accuracy} = 1 - \frac{\|y_i - \hat{y}\|_F}{\|y_i\|_F}
\]

(22)

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}
\]

(23)

\[
\text{Var} = 1 - \frac{\text{var}\{y_i - \hat{y}\}}{\text{var}\{y_i\}}
\]

(24)

\(\text{RMSE}\) and \(\text{MAE}\) are both measures of prediction error, with smaller values indicating better predictions and larger values indicating worse predictions. \(R^2\) and \(\text{Var}\) calculates the correlation coefficient to measure the ability of the prediction result to represent the actual data. The larger the value, the better the prediction effect.

4.3. Parameter Design

In this experiment, we choose the sym5 wavelet, the number of decomposition layers is 3, the learning rate of model parameters is set to 0.001, the batch size is set to 64, and
the training epoch is set to 3000. Because the prediction accuracy may be affected by different numbers of hidden units, we try different numbers of hidden units and compare the predictions to choose the best number of hidden units. The number of hidden units is selected from [16, 32, 64, 128] and the change in prediction accuracy is analyzed. As shown in Figure 10, the horizontal axis indicates the number of hidden units and the vertical axis indicates the change in metrics. When the number of hidden units is 64, the RMSE and MAE are the smallest. As the number of hidden units increases, the prediction accuracy increases first and then decreases. The main reason for this is that when the number of hidden units exceeds a certain level, the complexity and computational difficulty of the model increases greatly, leading to a decrease in prediction accuracy. Therefore, in all experiments, the number of hidden units is set to 64. Furthermore, the Adam optimizer is selected during the training process, and it will be used to calculate and update the network parameters of the model training and output so that the parameters are close to or reach the optimal values.

![Figure 10. Comparison of prediction performance under different hidden units.](image)

**4.4. Experimental Results and Analysis**

The experiment uses 12 historical data samples to predict the future 15-min, 30-min, and 45-min section speeds. Figure 11 shows the results of the visualization of the predicted future 15-min section speed using the WSTGCN model for four sections selected from the 16 sections. Black in the figure indicates the predicted section speed, red indicates the real section speed, the values marked by RMSE, MAE indicate the overall evaluation index of the section, and the orange and green bars indicate the RMSE and MAE calculated every three hours. It can be seen that the trend of the predicted speed and the actual speed are similar, and the variation between RMSE and MAE is small, indicating that the model can accurately predict traffic speed. From Figure 11c,d, it can be seen that the prediction accuracy will be reduced in the section where the traffic speed changes are complicated. As seen in the position of the red rectangle, the model can capture similarly varying trends in the face of sudden speed changes. To verify the reliability of the WSTGCN model, six baseline methods are used for comparison. These eight baseline methods are HA, SVR, ARIMA, GCN, GRU, LSTM, Spatial-Temporal Dynamic Network(STDN), which consists of CNN and LSTM, and GCN-GRU.
Table 3 shows the results of the evaluation metrics of the different models for predicting the section speed for the next 15 min. It can be seen that the results in all five evaluation indicators of WSTGCN are better than the baseline method. Compared with the GCN-GRU model, the RMSE of WSTGCN is 20.28% lower than that of GCN-GRU, and the MAE of WSTGCN is 11.69% lower than that of GCN-GRU. Because the GCN-GRU model does not consider the volatility of the data, which leads to lower prediction accuracy, and also proves that it is feasible to use WSTGCN to improve the accuracy of the prediction.

Compared with STDN, WSTGCN has 21.24% lower RMSE and 12.72% lower MAE, which indicates that the realistic expressway road network structure, GCN can capture spatial correlation better than CNN. The RMSE of WSTGCN is reduced by 55.68% and MAE is reduced by 59.13% compared with GCN, which shows that good prediction results cannot be obtained by considering only the spatial characteristics of expressway nodes without considering the temporal characteristics of their characteristic attributes. The RMSE of WSTGCN is reduced by 20.74%, 21.10%, and MAE is reduced by 13.18%, 14.01% compared with GRU and LSTM, which shows that ignoring the correlation between nodes among expressways when performing section speed prediction is also not achieving good prediction results. Therefore, combining GCN with GRU, which takes into account the spatial characteristics of nodes as well as the time series characteristics of nodes, can better improve the prediction results. Second, the RMSE of GRU is 79.91% lower and MAE is 53.72% lower than that of GCN, which is because GCN only considers the spatial characteristics of expressway nodes and does not consider the temporal characteristics of node feature vectors, and also indicates that the future section speed is more dependent on the section speed of historical time series. Compared with GRU, the RMSEs of HA, ARIMA, and SVR are about 46.87%, 54%, and 1.19% higher. Compared with WSTGCN, the RMSEs of HA, ARIMA, and SVR are higher by roughly 57.64%, 63.33%, and 21.22%, which is mainly caused by their poor nonlinear fitting ability to complex spatio-temporal data.
Table 3. Comparison of 15-min speed predictions.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>Accuracy</th>
<th>$R^2$</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>5.2682</td>
<td>2.9406</td>
<td>0.9457</td>
<td>0.4764</td>
<td>0.4764</td>
</tr>
<tr>
<td>ARIMA</td>
<td>6.0858</td>
<td>4.2199</td>
<td>0.9364</td>
<td>0.0025</td>
<td>0.0011</td>
</tr>
<tr>
<td>SVR</td>
<td>2.8328</td>
<td>1.4903</td>
<td>0.9708</td>
<td>0.8489</td>
<td>0.8491</td>
</tr>
<tr>
<td>LSTM</td>
<td>2.8248</td>
<td>1.4872</td>
<td>0.9708</td>
<td>0.8491</td>
<td>0.8491</td>
</tr>
<tr>
<td>GCN</td>
<td>5.0358</td>
<td>3.1289</td>
<td>0.9481</td>
<td>0.5243</td>
<td>0.5263</td>
</tr>
<tr>
<td>GRU</td>
<td>2.8158</td>
<td>1.4729</td>
<td>0.9708</td>
<td>0.8513</td>
<td>0.8493</td>
</tr>
<tr>
<td>STDN</td>
<td>2.8335</td>
<td>1.4652</td>
<td>0.9709</td>
<td>0.8518</td>
<td>0.8518</td>
</tr>
<tr>
<td>GCN-GRU</td>
<td>2.7992</td>
<td>1.4481</td>
<td>0.9711</td>
<td>0.8525</td>
<td>0.8525</td>
</tr>
<tr>
<td>WSTGCN</td>
<td>2.2316</td>
<td>1.2788</td>
<td>0.9771</td>
<td>0.9060</td>
<td>0.9063</td>
</tr>
</tbody>
</table>

Then, the prediction performance of WSTGCN and other models at different time intervals are further discussed, and different models are used to predict the future 30-min and 45-min section speeds, and their prediction performance is compared. Tables 4 and 5 compare the effects of different models on section speed prediction at different time intervals. For the 30-min versus 45-min speed prediction, the RMSE is reduced by 18.85% and 8.67% for WSTGCN compared with GCN-GRU, 22.34% and 19.21% for WSTGCN compared with GRU, and 45.66% and 33.19% for WSTGCN compared with GCN. It shows that the WSTGCN model is also able to capture the spatial and temporal characteristics of the section speed well in the case of long-term prediction.

Table 4. Comparison of 30-min speed predictions.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>Accuracy</th>
<th>$R^2$</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>5.4618</td>
<td>3.0503</td>
<td>0.9457</td>
<td>0.4379</td>
<td>0.4379</td>
</tr>
<tr>
<td>ARIMA</td>
<td>6.0858</td>
<td>4.2199</td>
<td>0.9364</td>
<td>0.0025</td>
<td>0.0011</td>
</tr>
<tr>
<td>SVR</td>
<td>4.1162</td>
<td>2.2221</td>
<td>0.9575</td>
<td>0.6821</td>
<td>0.6821</td>
</tr>
<tr>
<td>LSTM</td>
<td>3.7531</td>
<td>2.1942</td>
<td>0.9586</td>
<td>0.7106</td>
<td>0.7106</td>
</tr>
<tr>
<td>GCN</td>
<td>5.0358</td>
<td>3.1289</td>
<td>0.9481</td>
<td>0.5145</td>
<td>0.5166</td>
</tr>
<tr>
<td>GRU</td>
<td>3.5607</td>
<td>2.1305</td>
<td>0.9604</td>
<td>0.7258</td>
<td>0.7289</td>
</tr>
<tr>
<td>STDN</td>
<td>3.4516</td>
<td>1.9482</td>
<td>0.9638</td>
<td>0.7684</td>
<td>0.7692</td>
</tr>
<tr>
<td>GCN-GRU</td>
<td>3.4074</td>
<td>1.8117</td>
<td>0.9649</td>
<td>0.7816</td>
<td>0.7819</td>
</tr>
<tr>
<td>WSTGCN</td>
<td>2.7651</td>
<td>1.5906</td>
<td>0.9715</td>
<td>0.8559</td>
<td>0.8569</td>
</tr>
</tbody>
</table>

Table 5. Comparison of 45-min speed predictions.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>Accuracy</th>
<th>$R^2$</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>5.6406</td>
<td>3.1526</td>
<td>0.9418</td>
<td>0.4011</td>
<td>0.4011</td>
</tr>
<tr>
<td>ARIMA</td>
<td>6.0858</td>
<td>4.2199</td>
<td>0.9364</td>
<td>0.0025</td>
<td>0.0011</td>
</tr>
<tr>
<td>SVR</td>
<td>4.9536</td>
<td>2.7176</td>
<td>0.9489</td>
<td>0.5411</td>
<td>0.5411</td>
</tr>
<tr>
<td>LSTM</td>
<td>4.6462</td>
<td>2.6085</td>
<td>0.9556</td>
<td>0.6208</td>
<td>0.6221</td>
</tr>
<tr>
<td>GCN</td>
<td>5.2113</td>
<td>3.1912</td>
<td>0.9462</td>
<td>0.4912</td>
<td>0.4937</td>
</tr>
<tr>
<td>GRU</td>
<td>4.3096</td>
<td>2.5291</td>
<td>0.9502</td>
<td>0.6625</td>
<td>0.6284</td>
</tr>
<tr>
<td>STDN</td>
<td>4.1096</td>
<td>2.2354</td>
<td>0.9597</td>
<td>0.7146</td>
<td>0.7159</td>
</tr>
<tr>
<td>GCN-GRU</td>
<td>3.7873</td>
<td>2.0209</td>
<td>0.9609</td>
<td>0.7305</td>
<td>0.7311</td>
</tr>
<tr>
<td>WSTGCN</td>
<td>3.4818</td>
<td>2.0231</td>
<td>0.9641</td>
<td>0.7718</td>
<td>0.7736</td>
</tr>
</tbody>
</table>

Figure 12 shows the evaluation metrics of different models at 30 min and 45 min. We can see that WSTGCN still has lower RMSE and MAE compared to other models for the same length of time. As the prediction time increases, the accuracy of the models gradually decreases, and WSTGCN still has high accuracy at different lengths of time. Therefore, WSTGCN has good long-term prediction ability.
5. Conclusions

In this paper, an expressway speed prediction method based on wavelet transform and spatio-temporal graph convolutional network is proposed. First, the ETC gantry transaction data are matched with the topological data to construct the vehicle travel time data. Then the vehicle travel time outlier detection algorithm is used to eliminate the time anomalies of each section, and then the section speed data set is constructed. Finally, the section speed data and topological data are input into the WSTGCN model for training and learning, and are compared with various other models for analysis. The experimental results show that the prediction accuracy of the WSTGCN model in expressway speed prediction is significantly better than other methods, and it can accurately predict the section speed of the expressway. In addition, there are some shortcomings in this paper, i.e., other factors (e.g., weather conditions) that affect traffic speed are not considered. Next, introducing other data sources to improve more accurate section speed prediction is a work we plan to continue in the future.

Author Contributions: Conceptualization, F.Z. and Q.R.; methodology, F.Z.; software, Q.R.; validation, J.T., F.G. and S.H.; formal analysis, F.Z.; investigation, Q.R.; resources, J.T.; data curation, Q.R.; writing—original draft preparation, Q.R.; writing—review and editing, F.Z.; visualization, L.L.; supervision, J.W.; project administration, S.H.; funding acquisition, F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the National Natural Science Foundation of China (41971340), the Special Funds for the Central Government to Guide Local Scientific and Technological Development (2020L3014), the 2020 Fujian Province “the Belt and Road” Technology Innovation Platform (2020D002), the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian (GY-Z19113). Crosswise project (No.GY-H-21021).

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from Fujian Expressway Information Technology Co., Ltd. Furthermore, are available from the authors with the permission of Fujian Expressway Information Technology Co., Ltd.

Conflicts of Interest: The authors declare no conflict of interest.
References


