Journal of Network Intelligence ©2026 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **11**, Number **1**, February **2026**

Real Time Prediction of in Transit Vehicle Locations Based on the Electronic Toll Collection Data

Zi-Yang Lin[1,*]

[1]Fujian Key Laboratory of Automotive Electronics and Electric Drive,
Fujian University of Technology, Fuzhou 350118, China
lin19970119@126.com

Fu-Min Zou[2]

[2]School of Electronic, Electrical Engineering and Physics,
Fujian University of Technology, Fuzhou 350118, China
fmzou@fjut.edu.cn

Ze-Rong Hu[1]

[1]Fujian Key Laboratory of Automotive Electronics and Electric Drive,
Fujian University of Technology, Fuzhou 350118, China
595307290@qq.com

[*]Corresponding author: Zi-Yang Lin

Received July 11, 2024, revised April 30, 2025, accepted August 2, 2025.

Abstract. *The existing ETC gantries are laid at the expressway hubs, which are remote and distributed, and the accuracy of ETC gantries data generated from the interaction between vehicles and gantries is small, which cannot be used as the data support for vehicle-road cooperation. Therefore, this paper carries out the research of real-time prediction of expressway in-transit vehicle location based on the above work, uses the multi-stage decision idea of dynamic planning algorithm to effectively divide the vehicle trajectory, and realizes the in-transit vehicle location prediction within the ETC section of expressway through the preservation and update of vehicle status, and its accuracy rate is as high as 94.34%, compared with other basic models, the algorithm designed in this paper combined with real-time Compared with other basic models, the algorithm designed in this paper combined with the real-time number warehouse system shows superior performance and realizes the second-level update of vehicle position prediction, which provides data support and performance guarantee for the functional application of intelligent vehicle-road cooperation.*
Keywords: ETC, expressway, vehicle location prediction, Dynamic Programming

---

1. **Introduction.** The construction of highways is the foundation of China's economic and social development, and as a symbol of modern transportation, it has grown rapidly. As of 2023, the mileage of highways has reached 177000 kilometers, ranking first in the world. With the rapid development of highways, the construction of intelligent management is at the forefront. The intelligent management construction of highways requires continuous improvement of China's high-tech and big data interconnection technology, strengthening the data linkage between vehicles, roads, and people, and promoting the safe and efficient management and operation of highways. How to use the ETC toll system based on the Internet of Things and big data technology to achieve intelligent management has become a key issue that urgently needs to be solved.

Obviously, utilizing ETC transaction data to achieve intelligent management of highways requires consideration of potential risks in data management [1]. Moreover, the sparsity of the deployment of ETC gantry positions also poses challenges to the intelligent management of expressway.

In expressway management, ensuring vehicle safety and traffic stability is of utmost importance in achieving intelligent management [2]. In order to achieve real-time monitoring of vehicles based on the ETC system, continuous ETC gantries should be set up next to the highway lane. When vehicles pass through the ETC gantries, interaction data will be generated with these gantries, and the generated ETC transaction data can achieve full vehicle tracking. But as a result, the road section between the ETC gantry and the gantry is closed, making it impossible to monitor vehicle information within the section; And the distance between ETC gantry points is relatively long, with a scattered distribution. According to data from a certain province's highway, the longest ETC section exceeds 90000 meters in length, which may pose extremely high risks to vehicles; Due to practical reasons, the ETC gantry is expensive, and the physical installation of the gantry may affect changes in traffic conditions and economic losses. The sparsity of ETC gantry deployment and its internal black box characteristics greatly affect the intelligent management of expressway, bringing potential risks to driving vehicles.

In summary, in order to guarantee accurate prediction of expressway vehicle location and real-time recording of location data, we still face three major challenges as follows:

(1) The current vehicle location are based on GPS data and sensors for prediction, there are still problems such as signal instability and deviation in positioning.

(2) Facing the complex scenes of expressway, the signal interference degree is large, and the data upload is prone to drift.

(3) The vehicle position is uploaded to the control center for auxiliary decision-making, which requires a certain response time, thus leading to the time delay of vehicle position reporting.

Therefore, this paper adopts dynamic planning algorithm to realize the in-transit vehicle position prediction within the ETC section of the expressway, and also designs and combines a real-time number warehouse system to realize the second-level update of vehicle position prediction.

2. **Related work.** The key technology of intelligent vehicle road collaboration based on expressway, among which the fundamental construction is the research on vehicle position prediction [3], especially the real-time position of vehicles during driving. In navigation and early warning traffic safety applications, the most important basic data is the vehicle position [4]. Currently, the vehicle position is captured based on onboard sensors, and the onboard receiver obtains the vehicle position through signal interaction, and then makes predictions [5]. However, due to the numerous unfavorable factors such as tunnels and high mountains on expressway, it is easy to cause information interference, and the accuracy of onboard sensors may be damaged or even fail [6, 7]. Based on this, some researchers have started researching high-precision vehicle positioning based on LiDAR, but the current physical cost is too high to solve daily needs [8]. Therefore, low-cost, terrain free, and highly reliable location prediction are the conditions that this technology must meet [9]; In addition, it also needs to meet the demand for high-precision and real-time interaction 24/7. As an intelligent vehicle collaboration technology vigorously promoted by the country [10], ETC has high security protection, a large number of data dimensions, high-speed information exchange, and data advantages of long-distance traffic state perception [11].

Electronic Toll Collection (ETC) is an expressway toll system that utilizes dedicated short range communication for automatic toll collection. It can monitor vehicle conditions and automatically charge based on information such as travel distance and time. This system can greatly improve the efficiency and safety of expressway toll collection, alleviate traffic congestion and safety hazards caused by vehicle queuing and waiting for toll collection, while also saving vehicle travel time and fuel costs, improving driving efficiency and reducing transportation costs. As of 2022, the total number of ETC users in China has reached 285 million, covering most of the country's expressway. The popularization and promotion of the ETC system is of great significance for improving the efficiency and safety of road traffic in China, reducing the cost of vehicle traffic and energy consumption, and promoting the modernization and sustainable development of transportation.

The implementation of intelligent vehicle road collaboration technology based on the massive ETC system is one of the key issues of current world concern, among which the key issue is vehicle position prediction technology. It has a supporting role in over-the-horizon perception, navigation system, dangerous vehicle warning, etc. Min et al. [4] used a vehicle location prediction algorithm based on empirical modal decomposition and long and short-term memory network with simultaneous noise reduction, and the experiment proved that the prediction method improved the accuracy and stability of vehicle location prediction; the missing vehicle location seriously affected the data collection process of Telematics, Yu et al. [12] used the least squares algorithm of support vector machine to make up for the missing data, and the method predicted the vehicle location by combining the Kalman filter with the relational function; on the other hand, Peng et al. [13] used the vehicle location prediction model to generate the server offloading set and achieved the best allocation by the improved ant colony algorithm, which improved the task completion rate and resource utilization, Mei [14] oriented the vehicle networking system real-time monitoring of vehicle location information through mobile terminal and designed a probabilistic model for predicting vehicle location.

As known from above, not only the effectiveness of vehicle location prediction plays a fundamental role in expressway intelligent collaborative technology, but also its accuracy has been a difficult problem that hinders intelligent transportation. Facing complex road scenarios with a large degree of signal interference and vehicle position drift, Gao et al. [15] combined normal regression to identify vehicle drift data by setting dynamic thresholds related to vehicle speed and direction angle, and optimized joint filtering to eliminate position drift data; Alzyout and Alsmirat [16] used an automated ARIMA model for dynamic vehicle GPS position The model considers a specific window in the vehicle's historical position and uses this historical data to predict the next position, and experimentally verifies that their model greatly improves the accuracy of position prediction; improving the model is not the only solution to improve the accuracy of position prediction, Long et al. [17] predicted the position of a single vehicle by the regularity and preference of the vehicle, and their proposed model extends the LSTM with memory the LSTM to store all the output invisible states for vehicle prediction. Finally, Mei and Chen et al. [18, 19] also proposed relevant privacy protection schemes based on the operating environment of vehicles.

3. **Relevant Definitions.** The transaction information collected by the ETC gantry can reflect the vast majority of road traffic conditions on expressway in real time, and use ETC technology to remotely collect vehicle information, thereby providing better data support for intelligent vehicle assisted decision-making. However, due to the sparsity of the gantry position, the ETC gantry data presents a certain degree of dispersion, with high granularity, delayed vehicle information, and some models becoming complex. Therefore,

it is crucial to predict the location of vehicles within the ETC section. ETC vehicle data with accurate location information can more accurately and effectively improve the accuracy of intelligent vehicle road collaboration technology on expressway, simplify models, and improve model accuracy. The following are the relevant definitions for studying the location of vehicles in transit in this chapter:

Definition 3.1. Single vehicle section. In order to deeply explore the speed measurement information of the expressway without perception, this section further divides the single vehicle section $VSection$, as shown in formulas (1). $VSection$ is the single vehicle section, *preFlag* and *postFlag* are the front and rear gantry frames of the section, *preTime* and *postTime* are the transaction record time generated by the vehicle passing through the front and rear gantry frames, and the following *timedelta* and *speed* respectively represent the section travel time and average speed of the vehicle.

$$VSection = \{preFlag, postFlag, preTime, postTime, timedelta, Speed\} \tag{1}$$

Definition 3.2. Microelement section. This chapter refers to many small segments composed of segment trajectory points and front and rear gantry frames as micro element segments. Micro element segments are composed of adjacent gantry frames or segment trajectory points in longitude and latitude. The specific formula is as follows:

$$MiSection = \{PreMiNode, PostMiNode\} \tag{2}$$

Definition 3.3. Location of vehicles in transit. The location of vehicles on the expressway, Pos, consists of many longitude and latitude points, with two ETC gantry frames in front and behind. The longitude and latitude points in each of them form a time series of micro element sections, as shown in Figure 1. Each longitude and latitude point represents the possible location points that the vehicle may pass through. We define these continuous longitude and latitude points as the vehicle trajectories within the section, denoted as *Pos* and d, where x and y represent the longitude and latitude of the trajectory point, respectively. The formula is as follows:

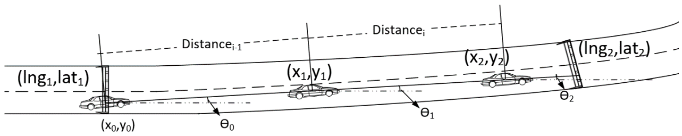

Figure 1. Location diagram of vehicles in transit.

$$Pos = \{d_1, d_2, d_3, ..., d_n\} \tag{3}$$

$$d_i = \{x_i, y_i\} \tag{4}$$

Definition 3.4. Section Data Flow. Expressway section data is propagated in Real-time data warehouse as a data stream, and we define it as *SectionDS*, as shown in formulas (5), where *id* is the marker of the data stream, *sym* is the aggregation field of the data stream, *bool* is a Boolean type variable, default to False, and $VSection$ is the single car section data $VSection$.

$$SectionDS = \{id, sym, bool, Vsection\} \tag{5}$$

4. **Real time prediction of in transit vehicle locations.** Vehicle position prediction is an algorithm research that predicts the position of vehicles in the next ETC segment based on their past trajectory. Its technology is mainly applied to the refined management of ETC vehicles, analyzing the regularity of ETC vehicle movement in the section and comprehensively evaluating the vehicle's driving position in the section.

At present, the ETC system collects vehicle position information through gantry data recording. These trajectory points record the spatial position of the vehicle at a certain time point, and the recorded data is ETC gantry data. During the ETC vehicle travel time, the vehicle's travel trajectory is composed of a series of trajectory points pieced together in chronological order, each trajectory point containing two elements: longitude and latitude. Therefore, when ETC vehicles have been driving for a period of time, the vehicle trajectory can be simulated as an ordered set P consisting of many trajectory points arranged in chronological order.

Generally speaking, an ordered set P is a continuous sequence with evenly distributed time intervals. However, according to statistics, in most cases, the time interval between two trajectory points in P ranges from 10 to 20 minutes. For high-speed vehicles, there are situations where the time interval is long and the degree of dispersion is large, which makes it difficult to effectively identify the distribution of vehicles and traffic conditions in the ETC section. Therefore, in order to refine the distribution points of vehicle trajectories, this article predicts the position of vehicle trajectory points in the ETC section, effectively improving position accuracy, and verifies it using "two passengers and one danger" data to ensure the effectiveness of the prediction.

4.1. **Data Pre-Processing.** This article will process the raw data. Due to possible anomalies in data format, data type, default values, and other aspects during the process, this section will analyze and process the integrity of the data in advance in sh. The specific process is as follows:

4.1.1. *Data format processing.* During the driving process, vehicles may encounter equipment failures, wireless signal crosstalk, and large vehicles obstructing communication, inevitably leading to the generation of topology. There is a format error in ETC, which cannot meet the data format specified by the function in the real-time data warehouse. For example, the *tradetime* field of ETC must be in the timestamp form of yyyy:MM:dd-H:M:S; The $shiftdate$ field must meet the formatting characteristics of the yyyy:MM:dd style; At the same time, Chinese characters may encounter garbled characters during device recognition and conversion. The above situation requires preparation before real-time data warehouse calculation.

4.1.2. *Data type processing.* When vehicles on expressway are driving at high speeds, due to excessive speed, the gantry frame is prone to generating default value data during interaction with vehicles. The existence of default values is directly related to numerical calculation, and we need to process the default values. For example, when the observer field is prone to default values when the vehicle is not recognized, we uniformly set the default value to 0, Facilitate data processing during the subsequent abnormal data analysis and processing stage.

4.1.3. *Default value processing.* When vehicles on expressway are driving at high speeds, due to excessive speed, the gantry frame is prone to generating default value data during interaction with vehicles. The existence of default values is directly related to numerical calculation, and we need to process the default values. For example, when the observer field is prone to default values when the vehicle is not recognized, we uniformly set the default value to 0, Facilitate data processing during the subsequent abnormal data analysis and processing stage.

Algorithm1 TWStream

Input: VSection.
Output: TVSection.
Step1: Init window = T_1, T_2, T_3, ..., T_n
Step2: **for** DS in window **do**:
Step3: PreTime, PosTime = getTime(window)
Step4: **for** SectionDS in T_1 **do**:
Step5: **if** sys = PreTime:
Step6: init AdjTable and init zip
Step7: sym, bool = [i for i in SectionDS]
Step8: **if** judge(bool)=True:
Step9: zip(index) = sym
Step10: AdjTable.insert(VSection)
Step11: bool = False
Step12: **else**:
Step13: find(sym)
Step14: AdjTable.insert(VSection)
Step15: **break**
Step16: TSection.append(AdjTable)
Step17: **End for**
Step18: **End for**
Step19: return TVSection
Step20: **End**

4.2. **Senseless real-time speed generation.** Cluster analysis is a key data mining task, which involves dividing observations into clusters, resulting in similar intra cluster observations and different inter cluster observations. In this article, we consider computing streams in the form of clustering, while the data stream, as a potentially infinite non-stationary sequence of continuous arrivals, is not feasible to achieve random access to the data, and storing all incoming data is impractical. We need to design a set of algorithms combined with time windows to aggregate the data stream to achieve the desired effect.

Based on the above analysis, this article proposes a data stream clustering Algorithm TWStream (Tumbling Window Tree). The TWStream algorithm uses labeling to modify the data stream. We use bool variables to cluster the data stream and label the stream. The value of the bool variable is defined as a Boolean type. If it has not been clustered before, the bool value is True, and after clustering, it automatically jumps to False. The data stream of aggregated field 3 and the data stream to be aggregated lack an aggregated stream, The bool variable is True, while all other aggregated bool values are False. The data stream begins to flow, and before aggregation, the bool variable of the data is tested, greatly reducing the search cost and avoiding the aggregation of a single data stream, as shown in Algorithm 1.

This article will use the senseless speed values generated by the TWStream algorithm to predict the position of vehicles in transit.

4.3. **Algorithm construction.**

4.3.1. *Framework of real-time prediction algorithm for vehicle position in transit.* In response to the demand for ETC big data in key technologies of intelligent vehicle road collaboration on expressway, this section designs a real-time vehicle position prediction algorithm framework based on ETC gantry data features through in-depth research on ETC gantry data features. This framework is based on real-time data warehouse, and is carried out based on a series of works in this paper, including Senseless real-time speed generation, as shown in Figure 2.

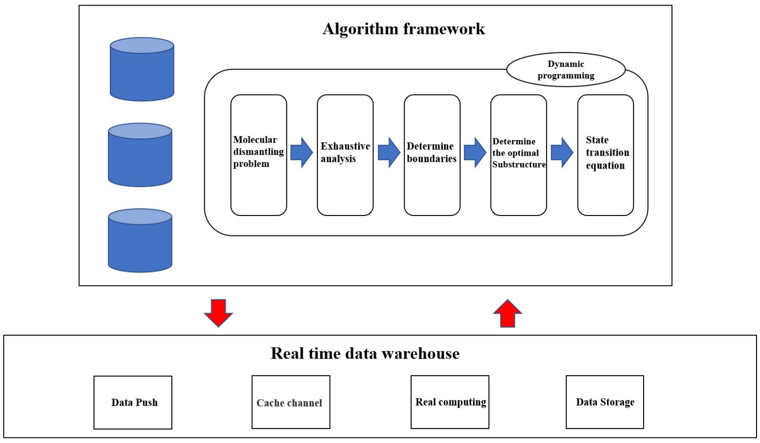

Figure 2. Framework of real-time prediction algorithm for vehicle position in transit.

This article designs a real-time prediction algorithm framework for expressway vehicle positions from five aspects, including the following five points:

(1) Splitting molecular problems: Splitting complex problems and simplifying their structure.

(2) Exhaustive analysis: Exhaustive problem structure, conceptualizing problem ideas.

(3) Determine boundaries: By using finite examples to demonstrate the regularity of sub problem solutions, determine the regular points as boundaries.

(4) Determine the optimal Substructure: the optimal Substructure is the condition that depends on the specific problem and the division method of the subproblem. Find out the rule and use the local optimal solution.

(5) State transition equation: Determine the current state based on the previous stage state and the results of the previous stage decision.

4.3.2. *Algorithm design based on dynamic programming.* The dynamic position of vehicles in transit studied in this chapter is estimated based on the ETC section. Considering environmental and road factors, the ETC gantry position is distributed discretely, mostly on sections with flat roads, long straight lines, high traffic flow, and reasonable toll station layout. However, the resulting length of ETC sections on expressway varies, and sections with long distances can easily affect the experiment. In order to avoid the influence of this factor on the experiment, this paper divides ETC vehicle driving sections based on dynamic programming.

According to the gantry and topology information provided by a certain province's expressway, it was discovered that there is a certain directional angle between the two ETC gantry frames. If the directional angle calculated between the front and rear gantry frames is directly used to estimate the vehicle position, the vehicle positioning may deviate from the expressway section, which affects the experimental results. Therefore, this paper uses the GaoDe API to crawl the track points of the section between the starting nodes

Algorithm2 Algorithm for predicting the position of vehicles in transit

Input: preFlag, postFlag, preTime, postTime, SeSpeed, distance.
Output: Pos.
Step1: x_i, y_i = getInfo(GNode)
Step2: $p_1, p_2, p_3, \dots, p_n$ = API(GaoDe)
Step3: $MiSection_i$, = divide(p_i)
Step4: init d_0 = preFlag
Step5: **for** d_i, MiSection in Section **do**:
Step6: θ_i = get(p_{i-1}, p_i)
Step7: $time_i = postTime_i$ - $preTime_i$
Step8: S_0 = dis($SeSpeed_i$, $time_i$, θ_i)
Step9: $d_i + 1 = d_i$ + change(S_0)
Step10: Pos.append(d_i)
Step11: return Pos
Step12: **End for**

of the section, uses the dynamic programming algorithm to divide the section, divides the section problem into the micro element section problem, and then solves the vehicle position in the micro element section. The density of section trajectory points has a great impact on vehicle position. The divided small section is less affected by azimuth angle and is less prone to position deviation, improving the accuracy of vehicle position. The position of vehicles at each moment has a significant impact on the prediction of vehicle position in the next infinitesimal section. The problem of vehicle position prediction in the ETC section of expressway can be decomposed into the optimal solution of multiple infinitesimal sections. The main steps are as follows:

(1) Stage division: For ETC segments with a length of distance, divide them into N discrete stages with a time length of t.

(2) Boundary determination: The initial position of the vehicle in the algorithm is set to the position of the front door frame of the section, and the algorithm ends when it reaches the next frame.

(3) State transition equation: The expressway ETC vehicles studied in this article default to average speed driving, and the state transition equation is:

$$\text{f(k+1)=}\begin{Bmatrix} preFlag, k = 0 \\ f(k) + change(distance), k > 0 \end{Bmatrix} \tag{6}$$

(4) Result comparison: Save the continuously updated vehicle status and compare the results. The specific algorithm is shown in Algorithm 2.

5. Research results and analysis.

This section will focus on the following aspects.

5.1. Pre-Introduction.

5.1.1. *Hardware.* To prevent single node collapse from affecting the integrity of the real-time data warehouse, this article adopts a cluster mode to build a real-time data warehouse system environment. According to the allocation of the laboratory, three servers were selected as the hardware environment for construction, and the specific configuration of each server is as follows, as shown in Table 1:

5.1.2. *Real time environment.* The real-time data warehouse system architecture consists of four modules: data acquisition module, offline data warehouse, real-time data

Table 1. Hardware

system hardware	*system configuration*
CPU	i9-10900K
Hard disk	4TB
Memory	64GB
operating system	CentOS 7.9

warehouse and data visualization. The main function of the data collection module is to continuously transmit real-time ETC raw data, including ETC gantry data, expressway topology data, ETC gantry information data, etc. The raw data is transmitted to MySQL for data storage and backup. Then, through components such as DataX and Maxwell, the data is transmitted to Hadoop cluster and Kafka cluster for intermediate storage. Finally, it is transmitted to the offline data warehouse and the ODS layer of the Real-time data Warehouse in the architecture for operation. The purpose of offline data warehouses is mainly to compare ETC offline data with ETC real-time data and timely correct parameters to ensure the accuracy and stability of the data produced by real-time data warehouses.

5.1.3. *Data Introduction.* There are three categories of experimental data. One is the intelligent transaction data generated by ETC gantry of Fujian Provincial expressway for 3 days from September 3st, 2020 to September 5th, 2020, as shown in Table 2, there are 54,474,098 pieces of transaction data, of which 4,611,924 pieces of trajectory data and 20,231,317 pieces of section data are fitted, and the other data is the topological relationship map between each gantry of the expressway and the distance between them. The "two passengers and one danger" data, it refers to buses with certain risks and road specialized vehicles for transporting dangerous goods. The data table collects and uploads GPS data through OBU devices on the vehicle, as shown in Table 3. All these data come from Fujian Expressway Information Technology Co., Ltd.

Table 2. ETC transaction data

Name	*Types*	*Examples*
PassID	String	01350119382305xxxxxxxx0501163019
Obuplate	String	xxxx9742
Entime	DateTime	2021-05-01 xx:xx:xx
Enstation	FixedString(4)	35xx
flagid	String	34xx01
tradetime	DateTime	2021-05-01 xx:xx:xx

5.2. **Data analysis.** This study selected expressway section prediction data and topology data from Fujian Province, China. To ensure the completeness of the data before the experiment, the following analysis was conducted in this chapter.

5.2.1. *Section distance.* Vehicle position prediction is based on the micro element segments after ETC segment division. Therefore, the length of ETC sections affects the division of micro element sections, and longer sections are prone to division deviation. Due to the large amount of section data, it is difficult to display. Therefore, this article selects the top 100 representative section lengths for display, with the longest section length exceeding 40 kilometers. According to statistics, 90% of the sections are within

Table 3. "Two passengers and one danger" data

Name	*Types*	*Examples*
Subtime	DATETIME	2020-09-05 18:23:07
Loctime	DATETIME	2020-09-05 18:23:04
Lat	FLOAT	119.554103
Lng	FLOAT	26.218683000000002
Speed	FLOAT	0
Obuplate	STRING	闽A0***D黄绿

20 kilometers in length, and only 9 sections exceed 40 kilometers in length, which falls within the reasonable range of section division, as shown in Figure 3.

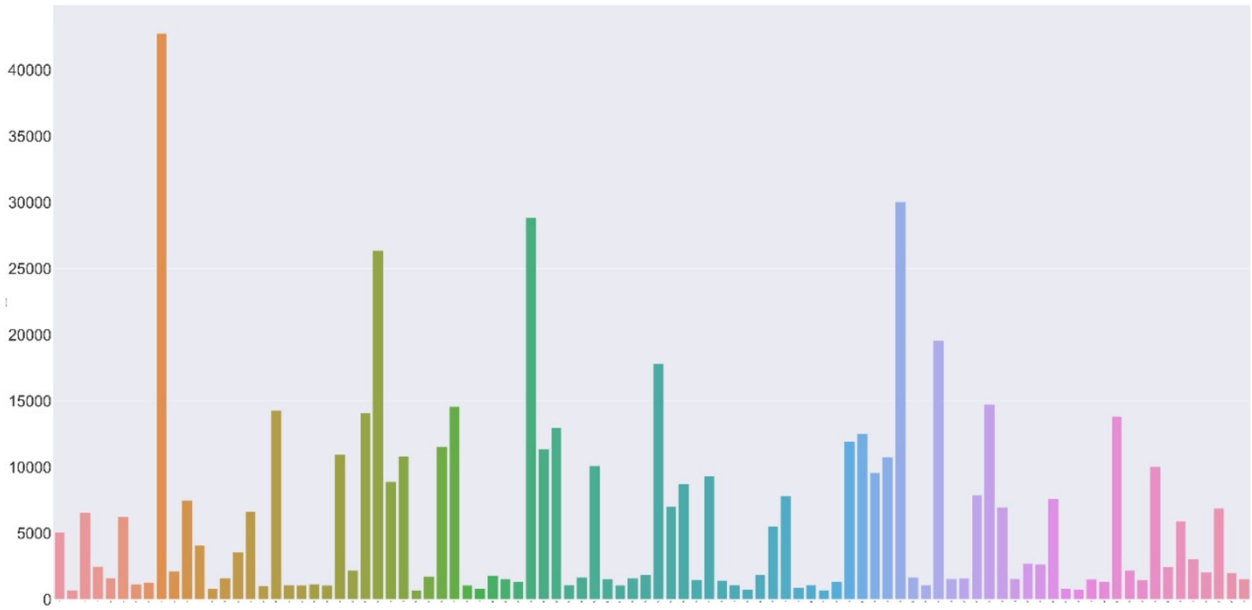

Figure 3. Schematic diagram of ETC gantry position.

5.2.2. *Section division accuracy.* The accuracy of vehicle position prediction depends on the number of position divisions. In this article, the GaoDe API is used to crawl the driving trajectory within the section. The accuracy of the crawled trajectory determines the accuracy of the section division. Therefore, the following analysis is conducted for the accuracy of each section division, as shown in Figure 4. In the ETC section, we divide the section every 100 meters on average, with a maximum division distance of nearly 250 meters, according to the evaluation based on the typical speed of 90 km/h for expressway vehicles, the travel time of 250 meters is 10 seconds, which is short and has a negligible impact on the steering angle.

5.3. **Deviation value of vehicle driving distance prediction.** Due to the difference between the predicted vehicle speed and the actual vehicle speed, the specific driving distance of the vehicle is prone to deviation. Within the expected time range, we calculated the distance traveled by the vehicle, assuming that the initial position of the vehicle is the starting gantry position O_j. According to the prediction algorithm, the final stopping position of the vehicle is D_i, while the actual stopping position of the vehicle is D_j. $\langle n_i, n_j \rangle$ represents the distance between the two points n_i, n_j, then we define the deviation of the vehicle's driving distance as distance_diff, the calculation method is:

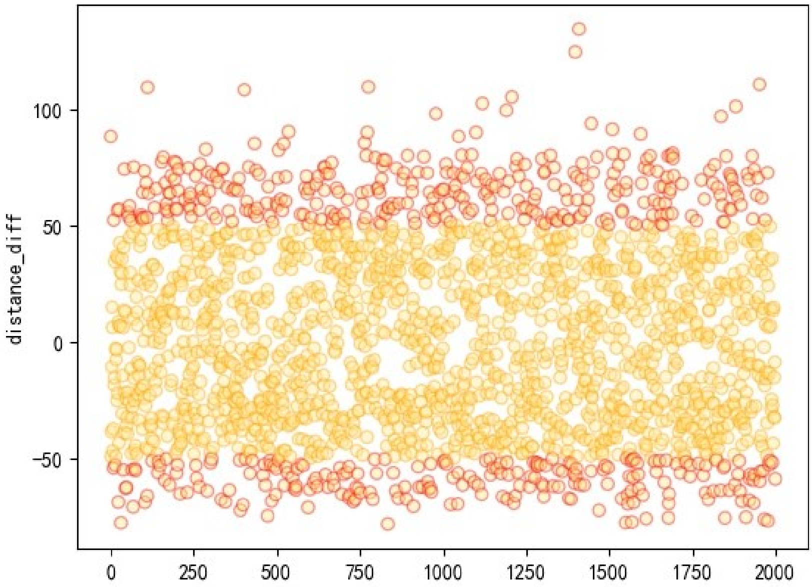

Figure 4. Schematic diagram of ETC gantry position.

$$distance\_diff = \langle O_j, D_j \rangle - \langle O_j, D_i \rangle \quad (7)$$

When comparing the final vehicle positions, we will set the distance_diff, The threshold of $diff$ is within -50 to 50m, which indicates that the final driving position of the vehicle is correct. If the value of distance_diff is not within the range, indicating that the prediction of the vehicle's driving position has failed.

As shown in Figure 5, the proportion of correctly predicted vehicle positions is shown, with most of the deviation in the driving distance of vehicle positions concentrated in the central area, resulting in relatively small errors.

Among them, the final driving position of vehicles was predicted for 4,601,298 section data, and the number of vehicles with correct positions reached 4,384,700, with a prediction accuracy rate of approximately 95.29%. Among them, the number of vehicles exceeding the current gantry position was 2091596, and the number of vehicles not arriving was 1,693,104.

5.4. **Traffic time after updating the speed of expressway sections.** Based on the analysis in the previous chapter, the relative error between the predicted and actual section speed values is less than 3%. However, for high-speed driving sections such as expressway, small speed differences can also be wirelessly amplified.

To reduce errors, we update the predicted speed values. When the vehicle enters the previous gantry, the real-time data warehouse starts running algorithms for distance accumulation and position prediction. When the next gantry transaction data is generated, the real-time data will update the speed based on the vehicle's travel time and section distance, replacing the predicted speed with the actual speed. The following figure shows the deviation between the predicted speed and the actual speed, with the predicted and updated values hovering around the range of -2 to 2.

5.5. **Vehicle position prediction effect.** This article uses "two passengers and one danger" data to verify the prediction results of vehicle positions in transit. The successful prediction coverage rate refers to the percentage of successfully predicted trajectory points in the section to the total trajectory points in the section. The threshold for successful trajectory point prediction is set to -30 to 30 meters, and the other ranges are trajectory

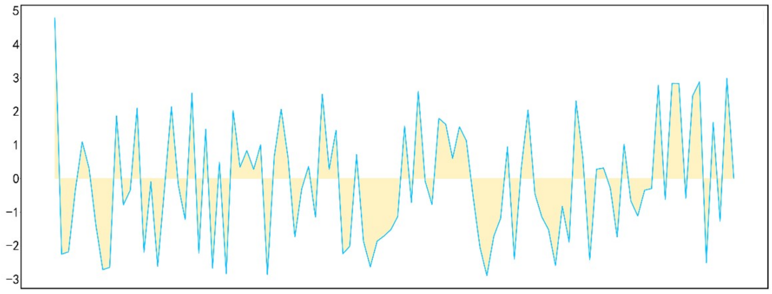

Figure 5. Schematic diagram of ETC gantry position.

points that failed prediction. Psr represents the predicted coverage of the segment, Nsuccess represents the number of successfully predicted trajectory points in the segment, and Nall represents the total number of trajectory points in the segment. Therefore, the for segment coverage is as follows:

$$Psr = \frac{N_{success}}{N_{all}} \tag{8}$$

This article conducts real-time estimation of the position of vehicles on the expressway. The experimental vehicle data volume is 4601298, and the prediction time of vehicle position is about 6.23 seconds for each data. The overall Psr of the data is 94.34%. This article shows the real-time prediction of the position of vehicles in continuous sections of traffic, as shown in Table 4. The minimum Psr is 88.82%, and the highest risk is 99.12%, demonstrating the reliability of the algorithm for predicting the position of vehicles on the expressway.

Table 4. Table for predicting the location of vehicles in transit

Plate	*VSection*	*Nsuccess*	*Psr*	*time(s)*
黑闽∗∗∗∗66	340131-340133	586	92.575039%	6.743
黑闽∗∗∗∗66	34012B-34012D	1149	92.961162%	12.912
黑闽∗∗∗∗66	340129-34012B	965	96.500000%	9.432
黑闽∗∗∗∗66	340125-340127	418	91.067538%	4.211
黑闽∗∗∗∗66	34011D-34011F	950	94.821428%	9.129
黑闽∗∗∗∗66	340117-340119	1298	93.122072%	13.118
黑闽∗∗∗∗66	340113-340115	1352	98.614150%	14.832
黑闽∗∗∗∗66	35071F-35071D	887	95.472527%	9.994
黑闽∗∗∗∗66	34010F- 340111	902	99.120879%	9.305

This article combines the real-time data warehouse system environment to achieve real-time prediction of the position of vehicles in transit. To highlight the superiority of the real-time data warehouse system and algorithm, this article compares the results with other basic models. The results demonstrate the superiority of the design environment and algorithm in this article, which can maintain advantages in accuracy while achieving second level prediction time, as shown in Table 5.

6. **Conclusions.** With the changing demand for intelligent vehicle road collaborative systems, road congestion and road safety have become a global social problem. The emergence of real-time vehicle position prediction technology has met the needs of intelligent vehicle road collaborative systems. However, nowadays, most of the real-time

Table 5. Comparison Table of Vehicle Location Prediction Models

Section	*Evaluation indicators*	*random forest*	*XGBoost*	*KNN*	*Proposed method*	*Time*
Section 1	R2	0.9423	0.9666	0.9037	0.9736	6.7435
	MAE	0.1135	0.0977	0.1364	0.0757	
	RMSE	0.0922	0.0673	0.1133	0.0528	
Section 2	R2	0.9385	0.9182	0.8991	0.9543	9.4327
	MAE	0.0877	0.0859	0.1099	0.0668	
	RMSE	0.0779	0.0936	0.0863	0.0593	
Section 3	R2	0.9523	0.9182	0.9222	0.9619	9.3058
	MAE	0.1021	0.0824	0.0921	0.0783	
	RMSE	0.0793	0.0765	0.0832	0.0752	
Section 4	R2	0.9212	0.9537	0.9235	0.9641	4.2113
	MAE	0.1023	0.1153	0.0984	0.0915	
	RMSE	0.9532	0.8821	0.7941	0.0783	

vehicle position prediction technology relies on vehicle sensors and surveillance videos for position capture. Its high cost and susceptibility to terrain impact pose obstacles to the effective integration of roads, vehicles, and related elements in intelligent vehicle road collaborative systems. The emergence of the expressway ETC system compensates for the high cost and susceptibility to interference of vehicle sensors and surveillance videos; Meanwhile, most of the existing in transit vehicle location predictions use offline data warehouses, which contradicts the demand for real-time data interaction proposed in the 14th Five Year Plan.Therefore, this article conducted an in-depth analysis of ETC gantry data and carried out the following work in response to the real-time characteristics of vehicle position prediction:

(1) Research on real-time data warehouse modeling of expressway for ETC systems. Design real-time data warehouses for expressway for research purposes, and compare the advantages and disadvantages of big data components for demand analysis. The designed real-time data warehouse consists of the raw data layer, data access layer, data warehouse layer, data aggregation layer, and data application layer from bottom to top. On this basis, Flink is used as a real-time computing engine, Kafka is used as a distributed message queue, and finally the results are stored through Clickhouse.

(2) Research on real-time prediction algorithm for the position of vehicles in transit within the ETC section. Analyze the characteristics of traffic flow in ETC section, use the dynamic programming idea to split the driving track of expressway vehicles, save and update the vehicle status, and finally calculate the driving position of vehicles in different micro element sections, which provides powerful data support for the intelligent vehicle road collaboration technology of expressway.

The continuously improving ETC system sets a high standard for traffic management departments to achieve intelligent vehicle road collaboration on expressway.The improvement of reliable data resources has brought new opportunities and challenges for in-depth research on real-time prediction of the location of vehicles on highways. Although this article has conducted a series of analysis, discussion, and research on real-time data warehouse systems for predicting the location of vehicles on highways. However, due to factors such as the time and conditions of graduate studies, as well as limited personal academic abilities, there are still some shortcomings and issues that need to be further explored in the above research.

(1) The driving speed of vehicles in the ETC section may be affected by factors such as congestion, service areas, and traffic accidents.The impact causes uneven speed on the road. This article mainly focuses on real-time prediction of the position of free flow vehicle routes. If multiple factors are considered, the predicted vehicle route position may have a large amount of deviation. In the future, multi scenario classification discussions can be added to real-time data warehouse systems, and deep learning algorithms can be used to predict the probability of occurrence for various situations, combined with real-time vehicle position prediction algorithms for prediction.

(2) In the face of complex environments, the accuracy of real-time prediction of on road vehicle positions is not ideal. The main theme of this article is to predict the location based on the predicted average speed of the section, although the speed of vehicles on highways shows a highly gentle trend, there may also be slight changes. In the future, more classification considerations should be carried out to analyze the traffic flow characteristics of sections in complex environments for speed prediction.

Acknowledgment. This work is partially supported by the Renewable Energy Technology Research institution of Fujan University of Technology Ningde, China (Funding number:KY310338), the 2020 Fujian Province "Belt and Road" Technology Innovation Platform (Funding number:2020D002), the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian (Funding number:GY-Z19113), the Patent Grant project (Funding number:GY-Z18081, GY-Z19099, GY-Z20074), Horizontal projects (Funding number:GY-H-20077), Municipal level science and technology projects (Funding number:GY-Z-22006, GY-Z-220230).

The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] T.-Y. Wu, Z.-Y. Lee, L. Yang, and C.-M. Chen, "A Provably Secure Authentication and Key Exchange Protocol in Vehicular Ad Hoc Networks," *Security and Communication Networks*, vol. 2021, 9944460, 2021.

[2] T.-Y. Wu, X.-L. Guo, L. Yang, Q. Meng, and C.-M. Chen, "A lightweight authenticated key agreement protocol using fog nodes in Social Internet of vehicles," *Mobile Information Systems*, vol. 2021, 3277113, 2021.

[3] W.-H. Luo, B.-T. Dong, and Z.-S. Wang, "Research on Vehicle Location Algorithm Based on Vehicle Road Collaboration," *Journal of Southwest Jiaotong University*, vol. 53, no. 05, pp. 1072-1077+1086, 2018.

[4] H.-G. Min, Y.-K. Fang, X. Wu, Z.-G. Xu, and X.-M. Zhao, "Position Prediction Method Based on Empirical Mode Decomposition and Long Short-term Memory under Global Navigation Satellite System Outages," *China Journal of expressway and Transport*, vol. 34, no. 7, pp. 128-139, 2021.

[5] C. Zhang, "Vehicle location data processing and dynamic route guidance technology," *Information Technology*, vol. 44, no. 12, pp. 129-133+138, 2020.

[6] Yujun Yue, Na Qiu, and Zhiyang Jin, *Algorithm Based on Cooperative Vehicle Infrastructure Systems[J]*, Journal of Chongqing University of Technology(Natural Science), pp. 1-13, 2023

[7] C. Li, Y. Fu, FR. Yu, TH. Luan, and Y. Zhang, "Vehicle position correction: A vehicular blockchain networks-based GPS error sharing framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 898-912, 2020.

[8] R. R, C. J, and G. T, "Automotive LiDAR technology: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6282-6297, 2021.

[9] F.-M. Zou, F. Guo, S.-J. Luo, L.-C. Liao, N. Li, and Y. Xing, "Research and Design of ETC Simulation Platform for Expressway," *Journal of System Simulation*, pp. 1-17, 2023.

[10] W.-Y. Chen, H.-C. Zhao, P.-F. Liu, J. Fang, and H. Sun, "Vehicle Detection Algorithm Based on Improved YOLOv3," *Control and Decision*, pp. 1-9, 2023.

[11] Y. Li, M. Shao, L. Sun, X. Wang, and S. Song, "Research on Demand Price Elasticity Based on Expressway ETC Data: A Case Study of Shanghai, China," *Sustainability*, vol. 15, no. 5, pp. 4379, 2023.

[12] D.-X. Yu, H.-B. Guo, Y.-F. Mo, and L.-J. Chan, "Vehicle Position Missing Information Prediction Method Based on Least Squares Support Vector Machine ," *Journal of Municipal Technology*, vol. 39, no. 04, pp. 17-22, 2023.

[13] WB.-P. Peng, G. Wang, C. Song, and JH. Yan, "Predictive Offloading Decision Algorithm for High Mobility Vehicular Network Scenarios," *Journal of Beijing University of Posts and Telecommunications*, pp. 1-6, 2023.

[14] C.-F. Mei, "Research on the Vehicle Position Prediction Based on the Vehicle Wireless Terminal Data," *Beijing university of chemical technology*, 2019.

[15] C. Gao, D.-F. Chu, and S.-X. He, "Vehicle Position Tracking Based on Joint Kalman-Gaussian Filter," *Journal of Transport Information and Safety*, vol. 38, no. 01, pp. 76-83, 2020.

[16] Mohammad S. Alzyout, and Mohammad A. Alsmirat, "Performance of design options of automated ARIMA model construction for dynamic vehicle GPS location prediction," *Simulation Modelling Practice and Theory*, vol. 104, 102148, 2020.

[17] W.-C. Long, T. Li,Z. Xiao, D. Wang, R. Zhang, A.-C. Regan, and H.Y. Chen, "Location Prediction for Individual Vehicles via Exploiting Travel Regularity and Preference," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 05, 2022.

[18] Q. Mei, H. Xiong, Y.-C. Chen, and C.-M. Chen, "Blockchain-Enabled Privacy-Preserving Authentication Mechanism for Transportation CPS With Cloud-Edge Computing," *IEEE Transactions on Engineering Management*, 3159311, 2022.

[19] C.-M. Chen, Q.-K. Miao, S. Kumar, and T.-Y. Wu, "Privacy-preserving authentication scheme for digital twin-enabled autonomous vehicle environments," *Transactions on Emerging Telecommunications Technologies*, 2023.