Heliyon 9 (2023) e21532

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research and design of simulation and verification system of intelligent expressway based on ETC big data

Fumin Zou [a,b], Nan Li [a,b], Feng Guo [a,b,*], Qiqin Cai [a,c], Xinjian Cai [a]

[a] Renewable Energy Technology Research Institute, Fujian University of Technology, Ningde 352100, China
[b] Fujian Provincial Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, Fujian, China
[c] School of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China

ARTICLE INFO

Keywords:
Framework for simulation
Feature extraction
Spatio-temporal feature
Traffic flow control
Multi-task scheduling
ETC system

ABSTRACT

Electronic toll collection (ETC) system records a large number of travel trajectories of vehicles on expressways, and it has a great potential application value. However, the current simulation system mainly focuses on simulating the characteristics of traffic flow while ignoring the real-time flow conditions of the road is difficult to calculate and display quantitatively, and the overall optimization cost is also notably substantial. Currently, there is a lack of a simulation system tailored for the ETC environment, which addresses the challenge of real-time traffic flow computation and holistic optimization, fulfilling the requisites of pertinent research. According to the topological structure inherent to an actual provincial road network on expressways, this paper devises a framework for a simulation system that conforms to the current ETC environment. We solved the critical problem of generating simulation data in the simulation system by establishing a Feature Extraction Algorithm for spatio-temporal features derived from ETC transaction data (Edata). Then we put forward Traffic Control Strategy Algorithm in ETC simulation system, which can provide decision indicators for simulating the control of traffic flow of the expressway. At the same time, we optimized the improved Multi-Task Scheduling Algorithm (ETC_MTS) based on the application scenario of real-time parallelism of multi-task on expressways, which provides better execution performance compared with the current mainstream algorithms such as Shortest Job First Scheduling Algorithm (SJFS), Priority Scheduling Algorithm (Priority), First Come First Serve Scheduling Algorithm (FCFS) and Round Robin Scheduling Algorithm (RR).

1. Introduction

As the contradiction between the increasing mileage and the low operational efficiency of China's expressways has garnered heightened attention, informatization and intelligence become the major trend in the development of expressways. Notably, our country also encourages and supports the intelligent development of expressways in terms of policies [1]. A myriad of data corroborates the assertion that the ETC network has emerged as a pivotal digital infrastructure underpinning traffic operation and management in the contemporary digital epoch. However, quantitatively calculating and presenting real-time traffic conditions on expressways poses challenges, and the cost of optimizing ETC construction is also substantial in practical research and analysis. Therefore, some experiments and related studies have been conducted at home and abroad on the architecture of simulation

* Corresponding author.
E-mail address: mapli@fjut.edu.cn (F. Guo).

https://doi.org/10.1016/j.heliyon.2023.e21532
Received 2 May 2023; Received in revised form 4 October 2023; Accepted 23 October 2023
Available online 3 November 2023

verification system mainly using virtualization technology. However, these researches can not meet the demand for virtual test environment in ETC environment, so a set of simulation system based on ETC environment has great practical significance and application value to improve the level of management and service quality of expressway.

At present, scholars at home and abroad have carried out a series of research on the related topics of intelligent expressway based on these Edata, mainly including several aspects such as identification of the maximum speed limit of expressway [2], traffic flow prediction [3,4], travel time of the vehicle as well as time prediction [5,6], expressway service areas, ETC gantries and other related research [7–12]. These studies are mainly based on vehicle-road cooperation technology to analyze the basic parameters of vehicle driving behavior, roadway capacity, and traffic operation delays. These efforts have promoted the development of intelligent expressways to some degree. However, the following challenges are still faced: 1) The framework of the current simulation system is mainly designed and researched surrounding traditional traffic models and Intelligent Transportation Systems (ITS). However, ETC systems have their unique characteristics, such as high speeds, large-scale traffic flow, and complex traffic scenarios, which lead to the fact that current simulation systems are not applicable to expressways; 2) Efficiently and accurately mining and extracting crucial information from Edata remains a persistent challenge.; 3) Existing simulation systems have difficulties in simulating massive in-transit traffic flow conditions on expressways, resulting in the inability to optimize traffic flow control on expressways; 4) Task scheduling scenarios are not considered in simulation systems.

In order to tackle the challenges as mentioned earlier, this paper proposes a framework of multi-task real-time parallel simulation of ETC of intelligent expressway. It defines the related service components in the simulation system about the existing research results. The main contributions of this paper are as follows.

1. Based on the needs of the simulation system of Edata, we propose a framework of multi-task real-time parallel simulation;
2. We mine and analyze different dimensional information in Edata by establishing a feature extraction model, which provides the basis for generating ETC data for simulation;
3. To effectively simulate the massive traffic flow conditions on expressways, we propose a traffic flow control strategy algorithm to achieve optimal control of the real-time traffic flow of the road network.
4. To improve the execution performance of the task analysis of the simulation system, we construct a multi-task scheduling algorithm based on the minimum cost flow algorithm and the established cost function, which conforms to the traffic flow of the simulation system.

The remainder of this paper is organized as follows: Section 2 provides an exposition on traffic simulation softwares and the traditional research on the calibration of the parameters of the simulation model. Next, the related methods are presented in Section 3, including the overall architecture of the system, Feature Extraction Algorithm, Traffic Control Strategy Algorithm, and ETC_MTS for traffic flow. The introduction of data sources and the experimental analysis are carried out in Section 4. Finally, we give the conclusion of this paper and provide an outlook for future research work.

2. Related work

2.1. Traffic simulation softwares

Traffic simulation technology plays a vital role in urban road management and control. It not only effectively reduces the occurrence of traffic accidents but also efficiently addresses issues of traffic congestion. Moreover, it assists in managing the traffic flow of a significant number of vehicles on primary roads and optimizing the complexity of road network infrastructure [13]. Most studies and research on the architecture of simulation systems at home and abroad use virtual technology. The current virtual test tools mainly include vehicle simulation, communication simulation, traffic simulation and driver simulation [12]. These virtual test simulation tools usually utilize pre-set parameters in the system to simulate the state and trajectory of a virtual vehicle during actual operation. The traffic flow data in the system presents a relatively desirable state during the simulation run, with relatively simple rules of interaction between traffic participants and clear priorities for various traffic behaviors.

The traffic simulation software currently used in the market mainly includes Corsim [13], TransModeler [14], Vissim [15], Paramics [16], Aimsun [17], Cube [18], and SUMO [19]. Traffic simulation software is primarily categorized into two main types: macroscopic simulation and microscopic simulation, where macroscopic simulation software is mainly applied to describe the overall characteristics of the traffic system, represented by Cube and TransCAD; microscopic traffic simulation software focuses on analyzing the actual operating characteristics of vehicles, which can more realistically reflect the microscopic behavior of the vehicle on the road, such as following, overtaking and lane change, represented by Corsim, Vissim and Paramics. Corsim of the United States integrates all the functional advantages of NETSIM and FRESIM which can ensure that it has the technical ability to dynamically simulate and display the changes of the actual road conditions in real-time urban traffic and dynamic road traffic simulation. TransModeler of Caliper Company of the United States is a multi-functional traffic simulation software, which displays vehicle operation and traffic conditions through a GIS-T graphical interface. Vissim of PTV company in Germany is a product designed based on a traffic behavior model, which can be employed to simulate traffic conditions within diverse social contexts, facilitating intelligent simulation and analysis of real-time urban traffic control models. Paramics, originating from Britain, has gained widespread adoption in road traffic planning and strategic management research, which can not only simulate road traffic networks, but also facilitate vehicle control through vehicle-road coopertative simulation. Aimsun, a simulation software developed in Spain, is primarily utilized for urban and road simulations on expressways. The time step of the simulation scene is 1~10 s, which can provide various models, such as following, lane change and headway interval. In addition, the simulation system can also meet a series of particular environments in the complex scenarios of real-time simulation, including complex signal transmission control and traffic accident scene rescue. Cube, developed by Citilabs, stands as one of the most extensively employed professional software systems in the field of traffic system planning and analysis, both domestically and internationally, which includes more than ten types of control software, such as Cargo flow forecasting and prediction, Voyager passenger flow forecasting, and Avenue transportation planning and control software. SUMO is an open-source, microscopic, and multimodal traffic simulation software developed by Deutsches Zentrum für Luft-und Raumfahrt (DLR). It can be controlled individually for each vehicle and is commonly used for the development of traffic control models.

However, the above simulation software is mainly applied to the theoretical and empirical study of the parameters of the dynamic operation and control of the relevant systems in vehicle-road coordination, such as the ability of vehicle diversion control in section [15], the assessment of traffic collision [18], and the speed of vehicle traffic [19]. However, these software solutions cannot generate and simulate the optimized and improved system architecture based on the ETC environment, nor can they provide the traffic flow environment under the safety-assisted driving and interaction situations, which in turn leads to the effect of the virtual test often fails to meet the demand for the realism of the virtual test environment of the vehicle-road cooperative in the ETC environment.

2.2. *Traffic simulation models*

The parameter calibration of the vehicle-road coordination simulation model in the intelligent expressway environment aims to optimize the combination of parameters in the ETC environment by establishing a complete calibration framework and combining it with specific traffic simulation software. Currently, research on traffic simulation models, both domestically and internationally, primarily encompasses three crucial aspects: the model selection and parameter setting in the design phase, the scene selection and parameter setting in the use phase, and parameter calibration of the model.

Firstly, researchers delve into the characteristics of various simulation models and their suitability for specific scenarios during the model selection and design phase of parameter configuration. This is to meet the needs of specific projects and lay a solid foundation for subsequent simulation experiments. Burger et al. [20] considered the process of applying traffic simulation software based on simulation modeling and proposed the design phase refers to the choice between using macroscopic, mesoscopic and microscopic simulation models around the research problem, the level of detail of the research unit and the requirements of computational efficiency. Chao et al. [21] started with a discussion on three classes of traffic simulation models applied at different levels of detail. Then, they discussed how traffic simulations can benefit the training and testing of autonomous vehicles. Burghout et al. [22] presented a framework for implementing meso-micro hybrid models, offering a unified representation of traffic dynamics. Furthermore, the chapter presented a new loading method that demonstrated a superior performance as compared to existing approaches.

Secondly, the key in the stage of the model selection and parameter setting in the design phase is to use the existing information to check and verify the traffic simulation software based on the requirements of specific scenes, so as to effectively carry out simulation experiments. Shangguan et al. [23] proposed a new three-layer simulation test architecture based on the modeling of traffic subjects, the simulation of group behavior, and the analysis of test results. They conducted experiments on the control methods of different groups of intelligent decisions by selecting typical traffic scenarios such as intersections and road sections, which effectively improved the efficiency, scale, and coverage of mixed traffic simulation tests in vehicle-road coordination. Van Lint [24] proposed a generic multilevel microscopic traffic modeling and simulation framework that provides endogenous mechanisms for behavioral differences among and within drivers, and it can generate multiple plausible HF mechanisms to explain the same observable traffic phenomena and congestion patterns due to distractions. Huang et al. [25] constructed a C–V2X application scenario for ramp convergence and intersection passage, and established an evaluation simulation model for controlling vehicle-road cooperation in low and medium densities in traffic flow situations by modeling the scenario. Zheng et al. [26] proposed a simulation analysis and modeling method based on VISSIM for the impact of emergency traffic events on the traffic operation of expressways in mountainous areas. The proposed prediction model can predict 80 % of the maximum queue length variation with good engineering applicability. Mullakkal-Babu et al. [27] integrated lateral vehicle dynamics and yaw movements into a traffic simulation framework that has been used to simulate multi-lane traffic flows consisting of human-driven vehicles.

Finally, during the model parameter calibration stage, the complexity of the parameter space in traffic simulation poses a significant challenge in determining how to calibrate traffic model parameters based on available data. This process introduces a multitude of uncertainties, including factors such as heterogeneous driver behavior [28] and measurement data errors [29]. Many researchers have put forth various approaches for calibrating parameters in traffic simulation models, aiming to enhance the efficiency of model evaluation and attain more accurate simulation results. Yin et al. [30] proposed a multivariate distribution of traffic flow model parameters calibration method based on clustering based on the vehicle trajectory model, and the advantage of the proposed method is mainly in the description of the discrete nature of the turning traffic passage trajectory compared with the previous traffic flow parameters calibration methods. Ciuffo et al. [31] proposed a parameter calibration method for traffic simulation models with multi-step sensitivity analysis, which improved the efficiency of model evaluation by 80 %, considering the asymmetry in the degree of influence of the input parameters on the output results. Wang [32] established an analytical system for the calibration of VISSIM parameters of urban road traffic based on the principle of the orthogonal test method. A traffic simulation model was constructed using static traffic data gathered from an actual field survey, and the dynamic calibration of the parameters of the dynamic intersection model was realized. Zhang et al. [33] selected VISSIM and used the urban expressway interweaving area as a simulation case, with the following model and the lane change model, and applied the improved LH-OAT algorithm (ILH-OAT) for sensitivity analysis of the model parameters, and then applied GA to calibrate the key parameters, which made the simulation result closer to the real road traffic operation.

3. Methodology

3.1. The overall architecture design

The architecture design of the ETC data simulation system of intelligent expressway is divided into two parts, which are ETC Simulation Data Center and ETC Simulation Application System, as shown in Fig. 1. These two parts together form a comprehensive simulation platform to support the research and decision-making of intelligent expressways. ETC Simulation Data Center is responsible for responding to real-time traffic flow data on expressways and identifying features of real Edata through machine learning and other technologies. In the feature recognition process, toll booths, car models, road sections, and other factors are considered for modeling to generate simulated Edata suitable for each application scenario on the system. However, it is crucial to emphasize that the effectiveness and accuracy of the simulation system rely significantly on the quality of the Edata. If there are incomplete or abnormal situations in the collection of Edata, it will directly affect the authenticity and reliability of the simulation effect. In addition, ETC Simulation Data Center, as a data governance center, has the function of self-inspection and self-diagnosis, which is used to inspect and repair the abnormal data in Edata. This process holds immense significance for refining and optimizing the feature model algorithm, ultimately enhancing the accuracy and credibility of the generated simulation Edata.

In ETC Simulation Application System, although the system is capable of simulating the concurrent situations of multiple concurrency on expressways and providing support for multitasking debugging techniques (such as traffic flow and single-vehicle-in-transit supervision) at the micro level, it is worth pointing out that the current simulation system mainly focuses on simulating the characteristics of the traffic flow, while there is a limitation of the impact of the traffic incidents on the traffic flow. This limitation may affect the simulation and validation of response strategies for specific traffic events.

In order to solve these limitations, ETC Simulation Application System adopts three key algorithms: Feature Extraction Algorithm, Traffic Control Strategy Algorithm, and ETC_MTS for traffic flow, which are described as follows:

Firstly, Feature Extraction Algorithm plays a crucial role in the ETC Simulation Data Center to mine and analyze the information of different dimensions by establishing a feature extraction model. This provides the basis for generating simulated ETC data, which makes the simulated data more realistic and credible. However, Feature Extraction Algorithm itself faces the impact of Edata quality on its accuracy. The output of Feature Extraction Algorithm may be affected if there are incomplete or abnormalities in the input Edata.

Secondly, Traffic Flow Control Strategy Algorithms are introduced to achieve optimal control of the massive in-transit traffic flow conditions in the road network. This helps to better simulate the traffic flow control on expressways and makes the simulation results closer to the real situation. However, the implementation of these algorithms may demand a significant amount of computational

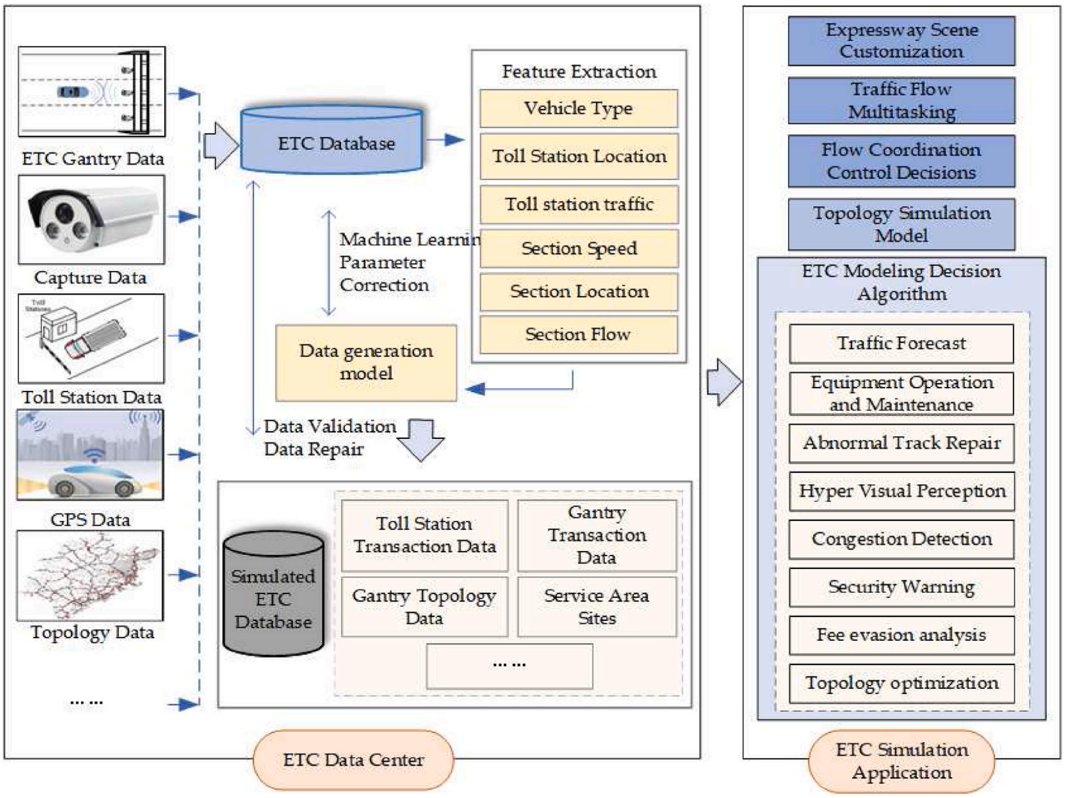

Fig. 1. Components of ETC simulation system of intelligent expressway.

resources, potentially posing challenges in environments with limited computational power due to resource constraints.

Finally, ETC_MTS is optimized to improve the execution performance of task analysis in the simulation system. Compared with the current mainstream algorithms, such as SJFS, Priority, FCFS and RR algorithms, the algorithm has better execution performance. It helps to process large amounts of simulation data more efficiently. However, the optimization of the algorithm may also have an impact on the complexity and resource requirements of the system, which needs to be weighed against its performance and resource overhead in real-world applications.

The following section will focus on several core algorithms of the ETC simulation system.

3.2. ETC feature extraction algorithm

Feature Extraction Algorithm based on Edata is the core algorithm of ETC Simulation Data Center on the simulation platform. The algorithm extracts the features of real Edata according to the platform's interface, including the extraction and analysis of vehicle type, driving behavior, space, time, and other features. The flow chart of the feature extraction algorithm is shown in Fig. 2. Considering the expansibility of the simulation platform, we have implemented the responsibility chain mode. In this mode, each independent feature recognition component can complete its feature extraction calculation work. Furthermore, processing components within the chain can seamlessly integrate into the chain sequence. Each processing component carries out its designated calculation and recognition tasks. Simultaneously, data beyond the scope of a particular component's capabilities is seamlessly passed on to the subsequent receiving component for further processing.

Edata include the passage information and toll information of vehicles, which can be used to reconstruct the trajectory of vehicles and estimate the position and speed of each vehicle through the fields of vehicle ID, toll station ID or gantry ID, and transaction time, so as to obtain the traffic operation situation of each section of the expressway. Edata have the characteristics of large scale and high density, and the data is composed of discrete points obtained by sampling, which has discrete and continuous mathematical characteristics. The algorithm uses the spatial and temporal distribution characteristics of traffic flow on expressways and its cyclical changes to establish a base feature library of cross-section flow, driving speed, driving preferences, and other basic features of traveling vehicles in different times and spaces on expressways. According to these basic information features library, the system can generate simulation transaction data in line with different scenarios, and meanwhile provide data guarantee for the research of intelligent driving assistance decision-making of ETC simulation system. In order to facilitate the research on feature recognition of Edata, we define Edata as follows:

Block: Two adjacent gantries on the expressway form a block.

$$Block = \{Node_1, Node_2\} \tag{1}$$

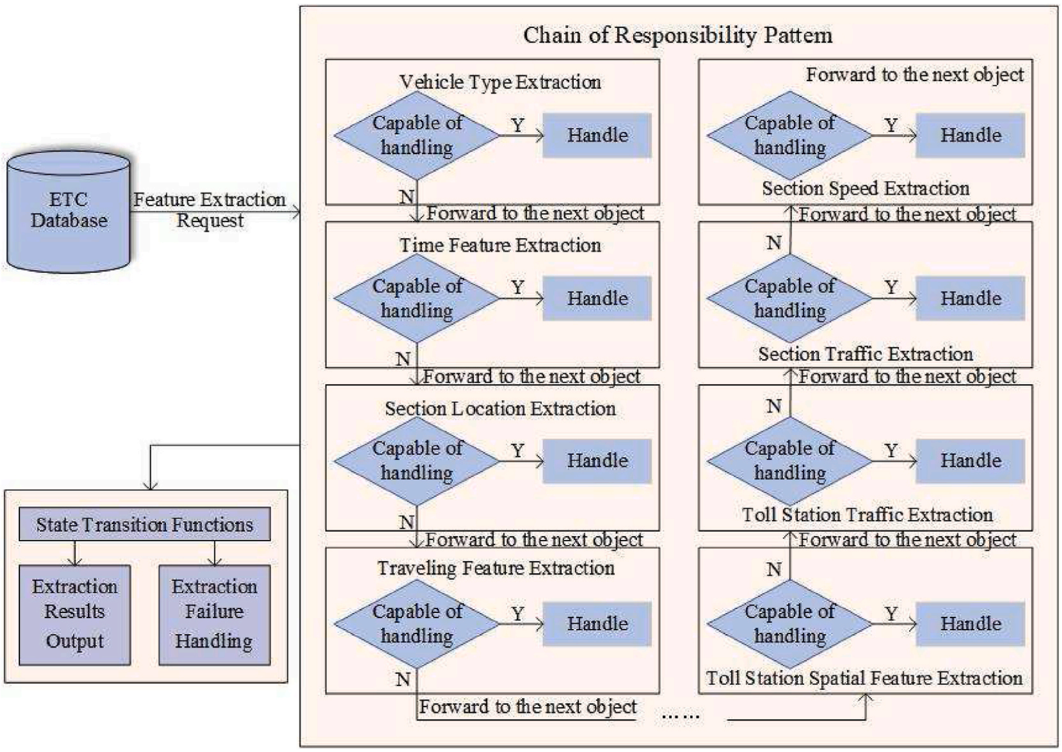

Fig. 2. Flowchart of feature extraction algorithm.

where $Node_1$ and $Node_2$ are two adjacent gantries in the road network on expressways. The schematic diagram of Block is shown in Fig. 3. The length of Block varies from a few kilometers to more than 10 km. The simulation platform can add or adjust the deployment of gantries in different scenarios according to the service requirements of the information from intelligent driving assistance decisions on the expressway.

Features of Edata (F): According to the trajectory of Edata and the gantry topology of passing places, we get independent and different features through functions of different dimensions. For instance, in the process of extracting speed features, flow features, and other attributes for different vehicle types, denoting the features as $F_1, F_2, F_3 \ldots F_n$, the following relationships can be deduced:

$$\sum_{i=1}^{n} P(F_i) = 1 \tag{2}$$

where $P(F_i)$ is the prior probability of the feature of Edata.

Constraints (C): The characteristics of Edata are closely related to the vehicles on the expressway in a particular environment, such as the number of vehicles on the roadway at that time and the maximum speed limit. We denote the relevant constraints as C and obtain the following equation.

$$P(F_i|C) = P(F_i) \times P(C|F_i) \tag{3}$$

On the basis of considering from the constraint C, we use $P(N_i|C)$ to denote the probability of occurrence of the situations that meet the constraint C, where N denotes the number of cases that satisfy the constraint C. The following equation can be derived.

$$P(N_i|C) = \frac{P(N_i|C)}{1 - P(N_i|C)} \tag{4}$$

Since $P(N_i|C)$ and $P(F_i)$ are equivalence relations, we combine Bayes' theorem and let $LS(N_i|C) = \frac{P(C|N_i)}{P(F_i|C)}$, we can derive the following.

$$P(N_i|C) = LS(C, N_i) \times P(N_i) \tag{5}$$

We add all the feature conditions that need to be extracted into the recognition component of the feature library for recognition and realize the sub-feature recognition of Edata by traversing the responsibility chain. The above example pertains to the extraction of sub-features, such as vehicle speed, specific to vehicle types within Edata. Feature Extraction Algorithm for Edata is implemented as Algorithm 1. The core part of this algorithm has three parts: the main program, the doTask program and the proceed program. The main program initiates feature extraction of Edata by registering the chain of responsibility. After loadRegisterFeatureParams reads the existing feature extraction components within the system, it proceeds to register these components within the environment through the register function. Following this registration process, the doTask program is executed to commence the preparatory work for feature analysis. During the doTask session, it initiates by employing the translate method to transform EData into spatio-temporal object data, referred to as featureData. Subsequently, it utilizes various FeatureAnalyzer components to perform feature extraction on featureData across different dimensions. The core part of the analyzer is proceeding, which filters the transaction data set containing the selected section according to the featureData, and then extracts the featureData according to three sets of parameters of the analyzer: type, vehClassSet, and timeSliceSet. The algorithm uses the mode of responsibility chain to make the simulation platform more flexible in extending the feature extraction components and reducing the coupling of data processing.

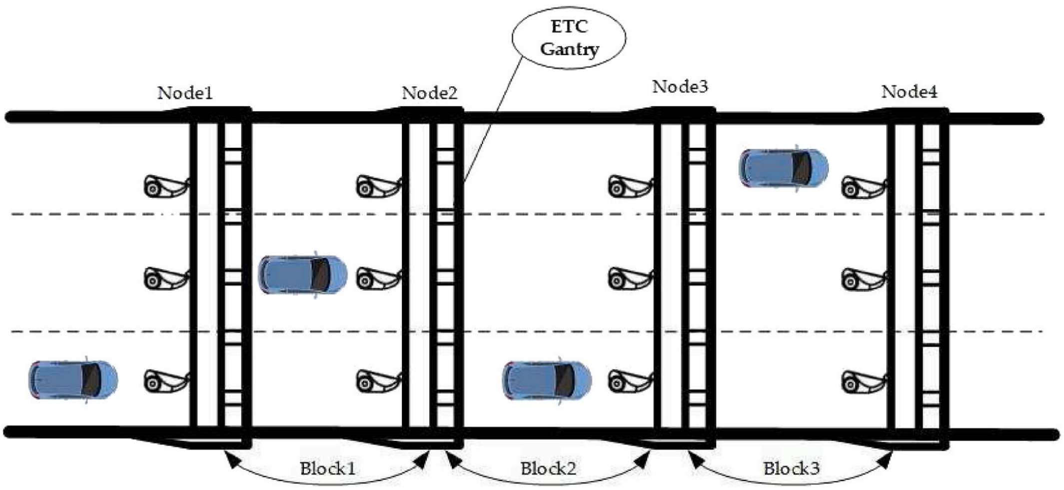

Fig. 3. Schematic diagram of Block.

Algorithm 1 Feature Extraction Algorithm for Edata

Input: ETC Transaction Data $Edata$;
Output: Characteristic Results R;
1: Chain chain = new Chain() //registering Chain of Responsibility;
2: featureComponents = loadRegisterFeatureParams() //loading the components of the system that have been asserted for feature extraction;
3: chain.register(featureComponents) //registering feature components with the responsibility chain;
4: R = chain.doTask($Edata$) // performing feature extraction;
5: Return R;
6: Def doTask($Edata$)
7: $FeatureData$ = translate($Edata$) // format conversion;
8: featureResultAndError = new FeatureResultAndError() //return object with extraction result and failure information;
9: **For each** featureComponents in FeatureAnalyzer **do**
10: FeatureAnalyzer analyzer = featureComponents.getAnalyzer() //getting feature components;
11: featureResultAndError.receive(analyzer.proceed($FeatureData$)) //specific feature extraction;
12: featureComponents.next()
13: **End For**
14: Return featureResultAndError
15: Def proceed($FeatureData$)
16: selected_data = getDataSet(featureData[Flagid]) //Filtering the transaction dataset containing the selected sections;
17: type = this.getAnalyzeType //getting the feature extraction type of this component;
18: vehClassSet = getVehClass(selected_data) //extracting the vehicle type characteristics of the transaction data set;
19: timeSliceSet = getTimeSlice(selected_data) //extracting the time characteristics of the transaction data set;
20: **For each** vehClass in vehClassSet **do**
21: **For each** timeSlice in timeSliceSet **do**
22: featureResult.append(getFeature(vehClass,timeSlice, type)) //according to vehicle type, time dimension, and other feature type parameters, data features are extracted;
23: **End For**
24: **End For**
25： Return R

3.3. Traffic flow control strategy algorithm

In the Traffic Flow Control Strategy Algorithm, the core objective of the simulation platform is to minimize the degree of congestion on the expressway to improve the overall traffic operation efficiency. In order to achieve this goal, the platform adopts the integrated traffic flow as the optimization index, and carries out dynamic scheduling and optimal control of the traffic flow on different road sections. Specifically, Traffic Flow Control Strategy Algorithm performs information-driven control calculations by obtaining information on various traffic flows generated by road sections on expressways in real-time during a specific period. During the calculation process, the platform performs dynamic coordination and optimization based on each section's traffic flow to reduce congestion and improve the capacity of sections. During the scheduling optimization process, the platform comprehensively considers the traffic flow situation at the diversion merging points on the expressway and identifies the toll stations and service areas that need to be coordinated and optimally controlled to minimize traffic bottlenecks and improve the road section's capacity and traffic efficiency.

The constraint relationship of road network traffic on expressways is as follows.

$$\Delta S = \sum_{i=1}^{M} \Delta F_{in}^{t} - \sum_{i=1}^{M} \Delta F_{out}^{t} + \sum_{i=1}^{N} \Delta F_{i}^{t} \tag{6}$$

where ΔS is the incremental traffic flow of the expressway; M is the number of toll stations; N is the number of Blocks; ΔF_{in}^{t} is the inflow from the toll station at time t, ΔF_{out}^{t} is the outflow from the toll station at time t, and ΔF_{i}^{t} is the flow rate at time t in the section. The cross-sectional flow rate of F is constrained as follows.

$$0 \leq F_{i}^{t} \leq F_{block\_mins}^{t} \tag{7}$$

where $F_{block\_mins}^{t}$ is the traffic degree of each Block at time t.

Because the expressway is a closed road scenario, when congestion occurs on a road section, vehicles can only drive away through the nearest toll station or wait until the flow rate returns to normal. Therefore, we need to optimize the flow control plan for each section of the simulated road network. The model of road network optimization is as follows.

$$maxF = \sum_{t=1}^{T} \sum_{k=1}^{N} \Delta F_{k}^{t} \Delta T \tag{8}$$

In Equation (8), $maxF$ is the increment of traffic in the optimized road network, N is the number of Blocks, T is the total number of periods during the scheduling period, ΔF_{k}^{t} is the increment of traffic at time t in the section, and ΔT is the duration of each period.

The algorithm controls the traffic at the associated intersection toll stations by finding the equilibrium value of F at time t.

With the control of the traffic flow control strategy module, the simulation platform runs as follows.

5. According to the traffic control demand on expressways, ETC data center generates operational data for the road network and control plans for traffic flow optimization every 15 min, based on the parameters configured for specific scenarios;
6. The system supervision module starts the heartbeat monitoring mechanism with every 15 min as a cycle and 96 times a day;
7. ETC data center generates traffic flow calculation model and in-transit vehicle generation plan operating parameters in the current 96 time periods of the road network;
8. The in-transit vehicle simulation module extracts vehicle characteristic data and related vehicle data, and calls traffic flow simulation calculation program to calculate and analyze in-transit traffic flow of the network on expressways.
9. ETC data center stores the traffic flow calculation results of the current scenario.
10. The in-transit vehicle simulation module adopts the current in-transit vehicle flow optimization control calculation model to carry out coordinated optimization calculation. This process generates a control strategy for the current road network traffic for a time step of 15 min.
11. The in-transit vehicle simulation generates the current coordinated optimization control strategy of vehicle flow and sends the relevant road section data to ETC data center for storage and display.

3.4. *Traffic multi-task scheduling algorithm*

In the ETC simulation system for intelligent expressways, simulating real-time multi-task scheduling of in-transit vehicles is a crucial challenge. The platform needs to process data from a large number of independent vehicles, calculate and update road conditions and vehicle status information in real time. At the same time, in order to keep the system performance stable, the expired data needs to be released in time. The current mainstream system architectures provide multi-thread scheduling algorithms to solve the throughput rate of the system through multiple concurrent operations, but the exclusive use of computing resources by multiple threaded tasks and the shared operation of data by multiple tasks will inevitably lead to the problem of data consistency. At the same time, the existing architectures generate thread locks when processing multiple tasks, which reduces the execution performance.

In order to meet the demand for simulation and parallel computation of a large number of vehicles for traffic simulation on expressways to achieve real-time updates of traffic flow and road state, this paper adopts concurrent execution to improve the computational efficiency of the system. The platform can more accurately simulate the concurrency of traffic flow on the expressway and realistically reflect the complex traffic situation through concurrent computation.

In order to optimize multi-task scheduling, this paper proposes ETC_MTS. ETC_MTS is based on the minimum cost flow algorithm and constructs a task cost function based on the task pool's processing time, thread utilization and waiting time of the tasks. This cost function is used to evaluate the priority of each task in order to better allocate computational resources to tasks with higher execution efficiency. With such a scheduling strategy, the algorithm effectively solves the problem of thread wait deadlocks during multi-task execution. It improves the utilization of the system's multi-threaded tasks, thus significantly improving the execution performance and efficiency of the simulation system.

In order to meet the system's goal of achieving multi-task and multi-threaded operation, we split the sections on expressways and supervised the overall Blocks according to the topology between the gantries. The system decomposes the road network into N Blocks based on the topology of the gantries. Whenever a vehicle travels into or out of a Block, that Block needs to identify, update, and compute the state of the section. Precisely, assuming that when a vehicle enters or exits a block, the system's kernel function will allocate a set of M computation tasks to different threads based on factors such as the vehicles within the block and the road conditions, including aspects like traffic flow, saturation, congestion, and average speed. These independent tasks all compete for the use of computational resources. The goal of task allocation is to effectively distribute the computational tasks triggered by these N in-transit vehicles to these M threads for execution to minimize the overall completion time. To achieve this goal, the platform employs a thread management structure, which is shown in Fig. 4.

In Fig. 4, Grid represents the storage area of the dynamic allocation management section, which adaptively allocates Block in the road network. The execution time of a computational task on each computational thread is determined by the prediction algorithm and the historical processing. *EET* is an $N \times M$ matrix of estimated execution times, and eet_{ij} in the matrix is the estimated execution time of computational task i on computational thread j.The execution time $R(T_j)$ of the computed thread T_j is the cumulative sum of the execution times of all tasks on that thread, it is specifically expressed as follows.

$$R(T_j) = \sum_{i=1}^{N} eet_{ij} \tag{9}$$

The completion time K of the task is specifically expressed as follows.

$$K = \max_{i=1}(R(T_i)) \tag{10}$$

The execution completion time E of tasks in the thread describes the running efficiency of submitted tasks of calculation, which are assigned to the best or nearly the best thread in the simulation system. The smaller the value of the execution completion time of tasks in the thread, the more tasks are assigned to run on the more desirable threads, which can be expressed as follows.

$$E = \sum_{i=1}^{M} R(T_i) \tag{11}$$

Load Balance (LB): LB is used to measure the balance degree of load pressure inside the thread pool. The higher the LB's stability, the better the state of LB in the pool, as defined below.

$$LB = 1 - \sum_{i=1}^{M} (K - R(T_i)) \,/\, (M \times K) \tag{12}$$

ETC_MTS is implemented as shown in Algorithm 2. ETC_MTS firstly calculates the processing time T_{proc}, waiting time T_{wait}, thread utilization variation V, task priority W_{task}, etc. Of all tasks by traversing eetList, secondly gets the resource matching matrix , load balancing matrix L, task fairness matrix F by regularization, then uses the matrix constructor buildMatrix to build job task matrix N to get the preliminary estimated task decision time set matrix S. Finally, the load balance and comprehensive matching in the set are judged according to the algorithm decision to determine whether the scheduling decision conflicts with the final thread task decision set D. If there is no conflict, the algorithm can add the task decision to the final thread task decision set D.

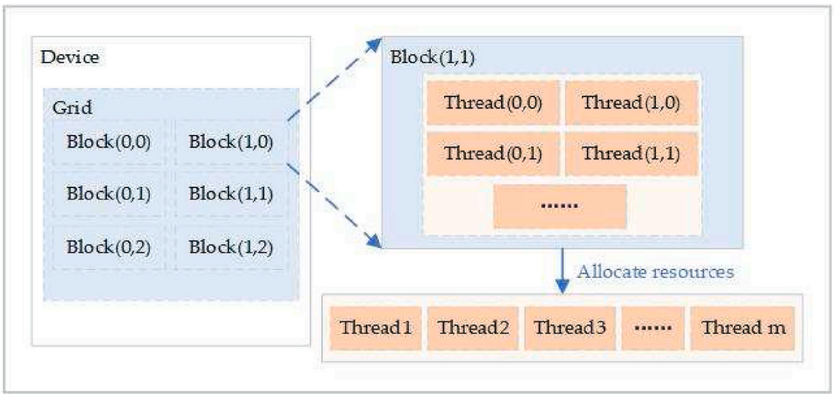

Fig. 4. The management structure of system's threads.

Algorithm 2 ETC_MTS Algorithm

Input: The set of tasks to be scheduled T; The set of schedulable threads M;
Output: The set of thread task assignment decisions D;
eetList = forecastTaskTime() //getting all currently available computing tasks;
For each eet in eetList **do**
    For each t in T **do**
        If eet can satisfy the condition of T**:**
            $T_{proc}(t, eet) = calcu(\frac{T_{qu}}{P_{proc} \times C_{proc}})$ //The processing time of the task is obtained by the weighted calculation of waiting time, processing process and overhead cost;
            $T_{wait}(t, eet)$ = The maximum value of waiting time in the set M of schedulable threads;
            $V = Variance\ (t, eet)$ // Thread utilization variance;
            $W_{task}(t, eet) = T_{submit}(t)$
        End If
    End For
End For
$R = 1 - Normalize(T_{proc} + T_{wait})$ //regularizing processing Processing time and waiting time;
$L = 1 - Normalize(V)$ //regularizing processing thread utilization;
$F = 1 - Normalize(W_{task})$//regularizing processing priority weights;
$I = R \times L \times F$
N =buildMatrix(I) //creating a running task matrix N;
S=initialDecisions(N) // The initial decision for the minimum cost flow is obtained by running the task matrix N;
For each $< t, eet >$ in C **do** //matching from maximum to minimum;
    If $< t, eet >$ does not conflict with the allocation decision set D **then;**
        D.append $< t, eet >$ //adding task decisions to the final decision set;
    Else:
        continue;
    End If
End For
Return D

Through this mechanism of task allocation and thread management, the system can effectively realize concurrent simulation and computation on expressways. Tasks within each Block can be executed in parallel among multiple threads. This fully utilizes the computational resources and improves the computational efficiency and throughput of the system. At the same time, the reasonable decomposition of topology and task allocation enable the system to accurately simulate and monitor the changes in traffic flow and status on expressways, which provides strong support for the technical research and application of intelligent expressway.

4. Results and discussion

4.1. Experimental settings and data description

In order to ensure the stability of real-time simulation and verification while meeting the demand for high-performance computing, this paper adopts the cluster mode to build the ETC simulation system. According to the laboratory allocation, we chose three servers as the hardware environment to build, of which the specific configuration of a single server is shown in Table 1.

Three servers work together in cluster mode. This can realize distributed computing and load balancing, thus ensuring the efficient and stable operation of the platform in large-scale traffic simulation scenarios. Cluster mode also enhances the reliability and fault tolerance of the system. Even if a node fails, other nodes can continue to provide services, ensuring the continuity and reliability of real-time simulation verification. ETC simulation system based on the cluster mode can better deal with the parallel computing needs of traffic data on expressways, real-time updates of road conditions and vehicle status information, and provide high-performance simulation results.

In addition, the simulation part of the experiment is based on the scenario laid out by 384 toll stations and 957 gantries in a province. The topology of toll stations and gantries on the expressway is shown in Fig. 5.

The dataset for the experiments in this paper is the Edata. Edata is obtained from more than 1000 gantries deployed in the whole road network on the expressway in a province. The collection period is from June 1 to June 5, 2021, and a total of 18, 726, 700 data are obtained, containing multiple information such as the identification of the vehicle, the time of the transaction, the number of the gantry, and the type of the vehicle after the transformation of the desensitization rules, partially shown in Table 2 below. Specifically, each transaction data contains all field information. The total number of vehicles in the data set is about 3.59 million.

4.2. Experimental analysis of ETC feature extraction algorithm

Edata contains the type of vehicles, the spatial characteristics of each toll station, the traffic volume of each toll station, the spatial characteristics of each road section and the speed characteristics. This experiment verifies and analyzes the traffic characteristics of vehicle type and time dimension, the traffic characteristics of each toll station space, and the speed distribution characteristics of vehicle type, respectively. In order to obtain the characteristics of the traffic flow on expressways in the time dimension, so as to accurately obtain the traffic flow evolution pattern by vehicle type, we used the dataset from June 1 to June 5, 2021 for our experiments. However, it should be made clear that the feature extraction conclusions obtained in this study are only applicable to the region where the data source is located, that is, the region of a southeastern province. Because of the driving habits of drivers, road network structure, and other factors, there may be differences in traffic flow characteristics in different regions.

The feature extraction experimental analysis is done based on a dataset of 384 toll stations and 957 gantry paving scenarios in a southeastern province. In this dataset, the statistics of the number of different types of vehicles are shown in Table 3. Among them, Class I of passenger vehicles accounted for a relatively large proportion, accounting for 76.54 % of the total number of vehicles, and Class I of trucks were the second largest, accounting for 10.92 %. Considering the different travel purposes of drivers, the vehicle types of mixed traffic flow have substantial uncertainty and randomness, and the traffic flow of different types of vehicles in the same toll station also shows different regularity. Therefore, it is difficult to accurately extract the flow characteristics of each toll station by only considering the mixed traffic flow at the toll station.

However, it should be noted that the obtained conclusions are only applicable to the region where the data source is located, and there may be differences in traffic flow characteristics in other regions. Therefore, Feature Extraction Algorithm should be applied cautiously in different regions and validated and analyzed in conjunction with local traffic data. This experiment divides the vehicle type features into three classes: passenger cars, trucks and work trucks, and extracts the traffic flow features in the time dimension based on these three dimensions for Edata. Edata is a kind of time series data that changes dynamically in continuous time, and the traffic flow shows regular changes within the same day. The traffic flow at each Block is counted by using 15 min as time slices through a time-based Feature Extraction Algorithm by vehicle type. A total of six different blocks are selected for the experiment to be analyzed, and these blocks have sufficient data samples to better reflect the trend of traffic flow corresponding to each time slice over time. The experimental results are shown in Fig. 6.

As can be seen from Fig. 6, there is a more significant percentage of traffic from passenger cars in each of the six blocks. Because the flow of passenger cars accounts for a large proportion of all traffic, resulting in a mixed traffic flow can only roughly reflect the trend of traffic flow changes in passenger cars. In contrast, the proportion of trucks and work trucks is relatively small, so it is difficult to extract features from the mixed traffic flow to obtain the travel pattern of each model, only by analyzing a single vehicle model can certain

Table 1
Configuration of the environment.

Systematic Hardware	System Configuration
CPU	i9-10900 K
Hard Disk	4 TB Mechanical Hard Drive
	512 GB Solid State Drive
Random Access Memory (RAM)	64 GB
Operating System	CentOS 7.9

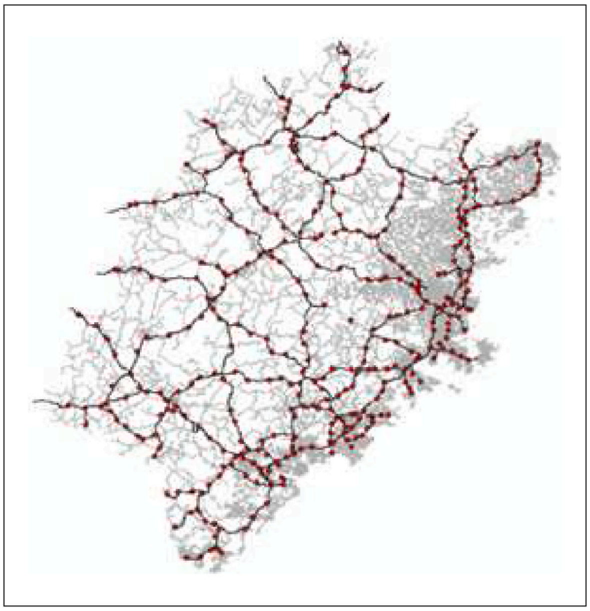

Fig. 5. The topology of toll stations and gantries on the expressway in a province.

Table 2
The attribute table of the field information contained in Edata.

No.	Field Name	Description	Example
1	VehID	Vehicle identification	S0***1(位)
2	TRADETIME	Transaction time	2021/6/4 00:00:00
3	FLAGID	Gantry ID	3502**
4	OBUID	Device MAC	66AD40**
5	ENTIME	Entrance time	2020/9/3 7:48:39
6	ENSTATION	Entrance toll station	46**
7	LNG	Longitude	118.56**
8	LAT	Latitude	24.85***
9	VEHCLASS	Vehicle type	1

Table 3
Statistics on the number of different types of vehicles.

Vehicle Type	Class	Code Name	Number	Percentage
Passenger Car	Class I	['1']	2,748,084	76.54 %
	Class II	['2']	24,372	0.68 %
	Class III	['3']	25,859	0.72 %
	Class IV	['4']	23,819	0.66 %
Truck	Class I	['11']	392,388	10.92 %
	Class II	['12']	79,685	2.22 %
	Class III	['13']	29,285	0.82 %
	Class IV	['14']	43,674	1.22 %
	Class V	['15']	19,448	0.54 %
	Class VI	['16']	195,834	5.45 %
Work Truck	Class I	['21']	3464	0.10 %
	Class II	['22']	3761	0.10 %
	Class III	['23']	331	0.01 %
	Class IV	['24']	210	0.00 %
	Class V	['25']	16	0.00 %
	Class VI	['26']	1	0.00 %

regularity be obtained. From Fig. 6, it can be seen that the daily traffic variation of each single vehicle has a certain regularity. Among them, the variation of the flow of passenger cars shows two clear peaks with higher flows. The traffic flow is low before 5:00. Traffic volumes proliferate after 5:00, and usher in the morning peak from 8:00 to about 10:00 a.m. The traffic flow starts to decline after 10:00, and shows periodic recurring changes from 11:45 to 18:30. The traffic flow decreases rapidly from about 18:30. The change in

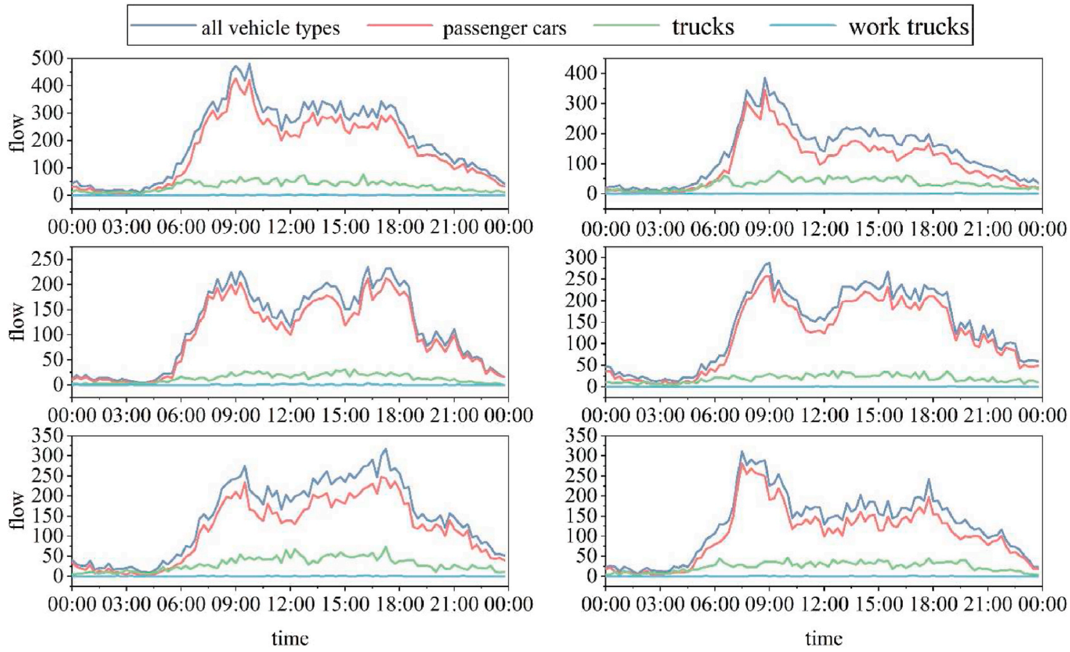

Fig. 6. Traffic flow characteristics based on time dimension by vehicle types in different Block.

the flow of trucks was relatively flat and the flow was small. Traffic has been increasing rapidly since 5:00, with a temporary peak at 6:30. Then the flow showed periodic recurrent changes, the traffic flow fluctuated constantly within a specific range, and after 17:30 the traffic flow was generally downward. The flow of work trucks changes exceptionally smoothly, with an average daily flow of about 10 vehicles, and the travel time is concentrated between 9:00 a.m. and 8:00 p.m. The experiments can identify the differences in traffic flow variation trends through the extraction of temporal features by vehicle type.

Due to the differences in the location of toll stations as well as their functions and drivers' travel preferences, the capacity of toll stations shows different regular characteristics according to the changes in the daily traffic flow of various vehicle types. The flow characteristics of the toll station will be closely related to the changes in the flow of vehicles, such as the existence of factories, living

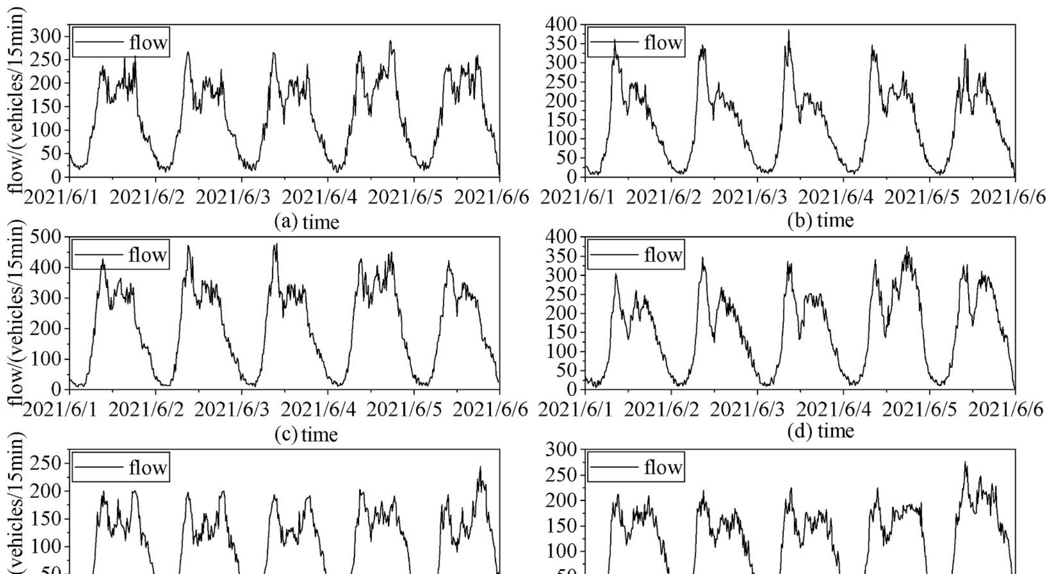

Fig. 7. Traffic characteristics of toll stations with different spatial types.

areas and scenic spots around the toll station have a particular impact on the travel of vehicles, mastering the relevant traffic flow patterns can provide the basis for expressway traffic control and prediction. In order to identify the traffic characteristics of toll stations, we extracted the traffic characteristics of ETC data set in a time step of 15 min. We prioritized toll stations located near factories, living areas and scenic spots, and selected from them those that showed significant advantages in terms of data volume. Finally, we analyze the traffic characteristics of six groups of toll stations with different spatial types, and the experimental results are shown in Fig. 7.

Based on the analysis results presented in Fig. 7, it is evident that the traffic flow characteristics of toll stations are influenced by their geographical surroundings. In Fig. 7, (a) and (b) belong to toll booths around factories, (c) and (d) belong to toll booths near living areas, and (e) and (f) belong to toll booths around scenic spots. We analyze them one by one below: The traffic flow at toll stations (**a**) and (**b**) peaked around 9:00 a.m., then slowly declined, and the traffic flow reached a small trough at noon, and the traffic flow fluctuated within a specific range in the afternoon and lasted for 5–7 h. The traffic starts to drop layer by layer at 17: 00 and 19: 00. The traffic flow at toll stations (**c**) and (**d**) began to decline after reaching the top at 9:00 a.m., the traffic flow reached the bottom at 12: 00, then the traffic flow began to rise, reaching the peak at 14 o'clock, and the traffic flow fluctuated within a certain range in the following hour, and then and fell in a "precipitous" manner. The traffic flow at toll stations (**e**) and (**f**) in non-working days is significantly more than in working days. On weekdays, the first traffic peak of the day is reached at 9 a.m. with about 199 vehicles. From 9 a.m. to 12 p.m., the traffic flow decreases slowly, then increases slightly and temporarily reaches a new peak at 2 p.m. The traffic flow decreases slightly from 2:00 to 4:00 p.m., then the traffic flow increases and reaches a new peak at 6:00 p.m. After that, the traffic flow gradually decreases to the bottom. In the non-working days, the traffic flow increased at all times, with a peak of 245 vehicles at 18:00 that day. This experiment provides a reference basis for the simulation of toll station traffic flow in the ETC simulation system feature library by realizing the extraction of the traffic flow features of each toll station.

There are also differences in the driving speed characteristics of different vehicle types in different sections. The experiment extracted and analyzed the driving speed characteristics of different models for the two blocks with significantly superior performance in traffic flow, and we sorted the experimental results in the order from smallest to largest and counted them at intervals of 5 km/h to obtain the histograms of the frequency distribution of speeds in different road sections. At the same time, in order to facilitate the comparison and analysis of the speed distribution characteristics of different models, we used Origin software to perform single-peak or double-peak fitting of Gaussian distribution. The R^2 values of the fitting results of all models are greater than 0.9, indicating that the fitting effect is quite reliable. The speed frequency statistics of some models of the two blocks and the corresponding fitting results are shown in Figs. 8 and 9.

From the analysis results in Figs. 8 and 9, it can be seen that the speed distribution curves of each vehicle model are generally distributed in a "mountain" shape in *Block* B. In contrast, some vehicle models are distributed in an "M" shape in *Block* A, which indicates that there are differences in the speed distributions of different vehicle models on different road sections, and that road conditions and vehicle performance have different impacts on the speed distribution. Road conditions and model performance have different effects on the speed distribution. The specific analysis is as follows.

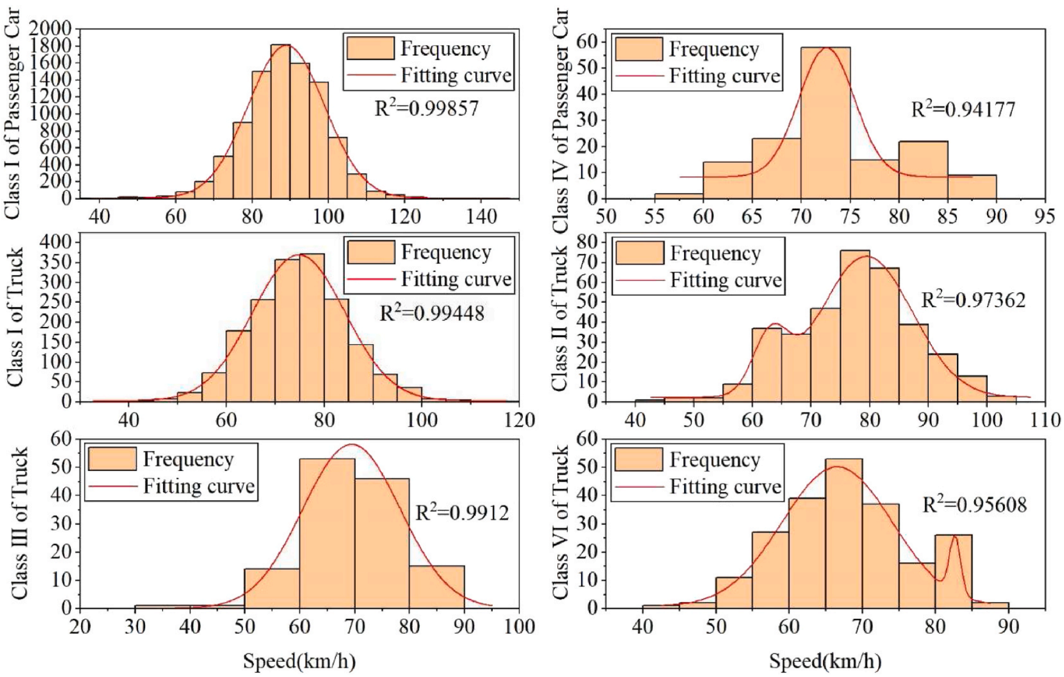

Fig. 8. Speed characteristics of some vehicle types in *Block* A.

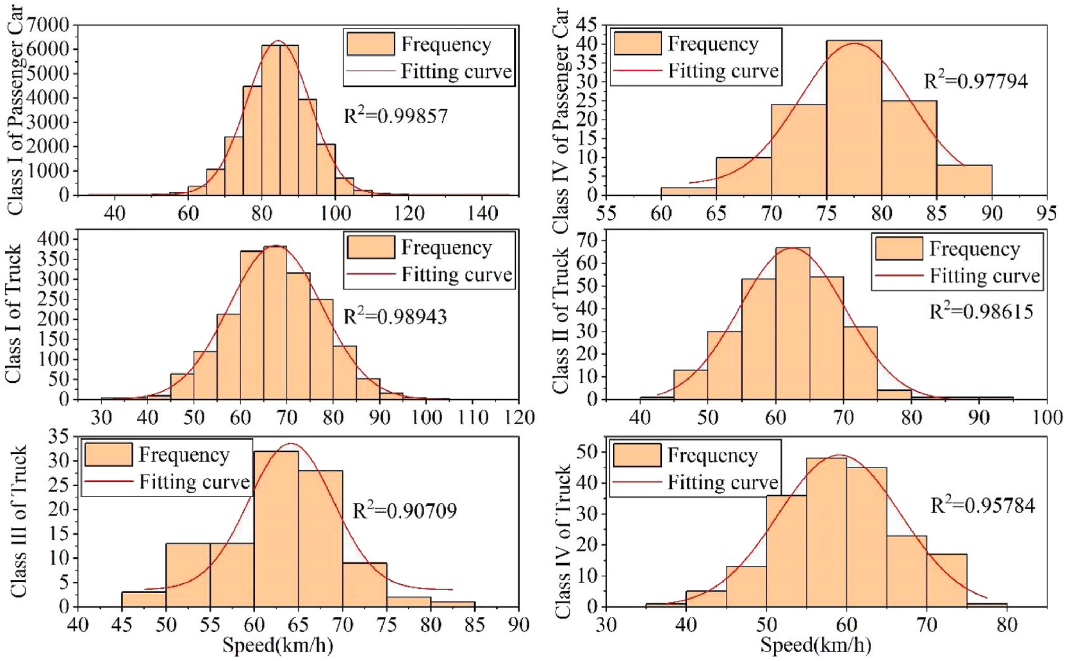

Fig. 9. Speed characteristics of some vehicle types in *Block* B.

1. The speed distribution of Class I of passenger cars in *Block* A and *Block* B are approximately normally distributed, and the maximum speed is concentrated at about 80 km/h;
2. The speed distribution of Class IV of passenger cars in *Block* A is more concentrated in (70,75], with a notable decrease in frequencies below or above this interval. Conversely, in Block B, the speed distribution is more focused around the (75, 80] range, and the speed distribution of Class IV of passenger cars in *Block* B is close to the standard normal distribution, while there is a significant decrease in the frequency of speeds below or above the speed interval in *Block* A;
3. The speed distribution of Class I of trucks in *Block* A and *Block* B are approximately standard. The driving speed of Class I of trucks in *Block* A is affected by the road conditions making more vehicles drive at low speeds.

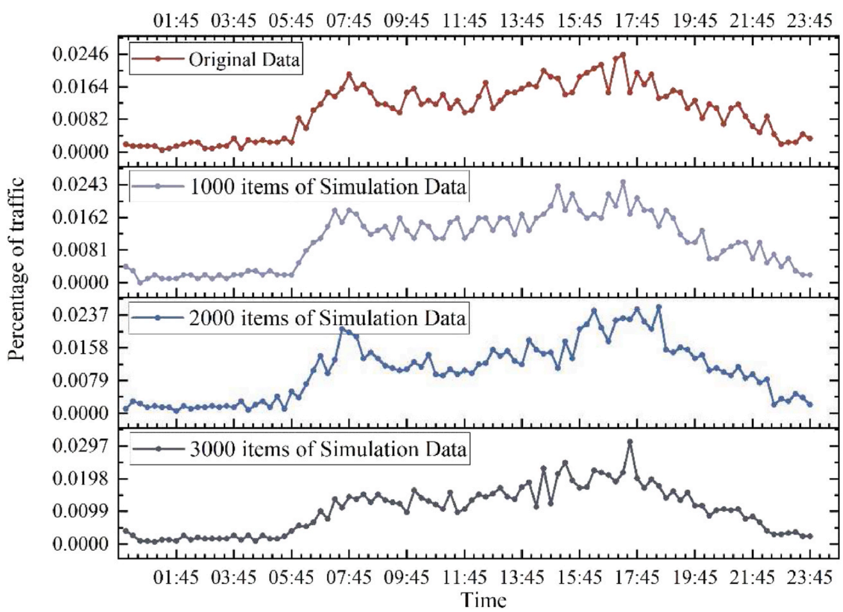

Fig. 10. Percentage of traffic flow for different orders of magnitude of simulation data versus actual traffic data.

4. The speed distribution of Class II of trucks in *Block* A and *Block* B have an enormous difference. As can be seen from Fig. 8, the speed distribution of Class II of trucks in *Block* A is hump-shaped, and the frequency reaches the maximum near 65 km/h and 80 km/h. In contrast, the speed distribution closely approximates a normal distribution in Fig. 9. Moreover, there are more vehicles in *Block* A driving at higher speeds.

Through the feature extraction of Edata, such as traffic characteristics of vehicle types and time dimensions, traffic characteristics of each toll station space, and speed distribution characteristics of vehicle types, the characteristic parameters of Edata are mined, which provides a reference for simulation data generation of the simulation platform.

In order to verify the accuracy and authenticity of the generated dataset in the simulation system, the experiment analyzes and compares one day's Edata of key road sections in Fuxia, generated by the system, with the actual traffic condition data. In order to more comprehensively assess the similarity between the simulation data and the real data, different orders of magnitude of simulation data were generated, including 1000, 2000 and 3000 data. In this paper, we summarize the previous research results [34,35], and comprehensively compare and analyze the three aspects of traffic flow, vehicle speed distribution and vehicle model distribution to verify that the simulation data can effectively reflect the actual traffic conditions. The comparison results are shown below.

(1) Comparison of Traffic Flow

The percentage of traffic flow for different orders of magnitude of simulation data compared to actual traffic data is compared in Fig. 10. The results show that both the real data and the generated simulation data of different orders of magnitude have a relatively low traffic flow percentage before 6:00 a.m., thus presenting the characteristic of lighter traffic load; the traffic flow percentage starts to increase gradually from 6:00 a.m. and reaches the first short-time peak of the day at about 7:45 a.m. In the following period, the traffic flow percentage continued to fluctuate slightly. Until about 5:30 p.m., the traffic flow reaches the highest peak value of the day. It then gradually decreases, thus indicating that the generated simulation data matches well with the real data. Simultaneously, the experiment conducted a Pearson correlation coefficient analysis between actual and simulated data traffic flow proportions. The calculated correlation coefficients for 1000 simulated data, 2000 simulated data, and 3000 simulated data compared with actual data are 0.915809, 0.907352, and 0.889066, respectively. These series of correlation coefficient results indicate high similarity in the traffic flow proportion trends between the generated simulated data and the actual data.

(2) Comparison of vehicle speed distribution

In order to compare the speed distribution characteristics of different orders of magnitude of simulated and actual data, we plotted the frequency histograms of speed distributions for the original data set and the simulated data set with different amounts of data. By fitting the histograms in Origin software using Gaussian distribution, we obtained R^2 values greater than 0.95 for both the simulated and original data fits. These histograms show that the frequency distributions of vehicle speeds are all generally distributed, as shown

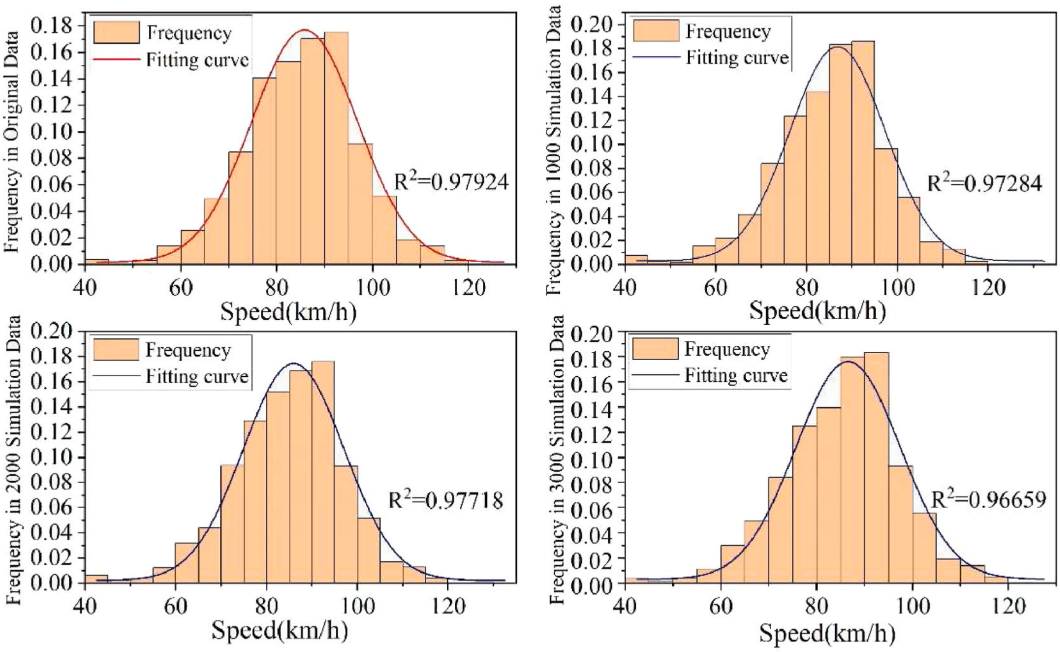

Fig. 11. Comparison of speed distribution characteristics between simulated and original data of different orders of magnitude.

in Fig. 11. The first plot shows the frequency distribution of speeds for the original data, where more than 80 % of the vehicle speeds fall within the range [70, 100]. Similarly, the generated simulation data also shows this trend. Specifically, it can be observed that the frequency of the speed distribution of the simulated data is consistent with that of the real data: the highest frequency occurs in the range of [90, 95], followed by [85, 90], and the third highest frequency occurs in the range of [80, 85]. This similarity further validates the feasibility of our simulation data to match the original data.

(3) Comparison of Vehicle Type Distribution

The distribution of vehicle types for different orders of magnitude of the simulation data compared to the original data is presented in Fig. 12. In each different order of magnitude, we can observe a similar trend, in which the highest percentage of vehicle models is still ['1'], representing Class I of passenger cars, with a stable percentage of approximately 79 %, and the percentage of simulated data changes minimal from the original data, with a range of −1.421 %–0.564 %; The following highest proportion is represented by ['11'], corresponding to Class I of trucks, accounting for around 9 %. Its share shows marginal variation from the original data, ranging from 0.05 % to 0.758 %. Then, ['16'], representing Class VI of trucks, ranks third in terms of share, at approximately 7 %. Its proportion undergoes slight changes compared to the original data, within a range of −0.85 %–0.291 %, essentially maintaining similarity to the original data share. The changes in the percentages of other vehicle types at different values are all relatively small, with the enormous difference being that the percentage of ['13'], which is around 0.331 % higher compared to the original data when the data volume is at 3000. The impact of smaller sample sizes may influence these variations. Generally speaking, when analyzing the percentage of models under the values of 1000, 2000 and 3000 compared with the original data, the trend of the percentage change of most models under different values is similar to that of the original data.

In summary, the simulation data of different orders of magnitude show similarity with the original data in the three aspects of traffic flow, vehicle speed distribution and vehicle model distribution. The experimental results are as expected, further highlighting the effectiveness of the simulation system.

4.3. Experimental analysis of Traffic Control Strategy Algorithm

This experiment uses 384 entrance and exit toll stations in a province to form 147,072 driving OD paths for traffic flow control strategy simulation experiments. The experiment aims to help managers optimize overall traffic flow, reduce the number of congested roadways, and improve the overall efficiency of traffic flow. A certain number of vehicles are randomly driven at the entrance of each toll station in the province, and the traffic flow data of a particular time slice of the province's expressway is randomly generated according to the vehicle type characteristics. According to Traffic Control Strategy algorithm, we identified the high-loaded road sections, that is, the road sections that are congested due to excessive flow, and the results are shown in Fig. 13.

It shows the road sections in a province that need to handle high load traffic flow at a particular period in Fig. 13. In the figure, we can see the expressway network of the province as well as the key bolded road sections that are identified as high load road sections. These high-load road sections have high traffic flow and need to be optimally controlled to relieve congestion.

The experiment generates one day's simulation data through the system and uses Feature Extraction Algorithm to obtain the average speed, section flow and traffic density of each section every 15 min, and identifies the congested road sections according to Traffic Flow Control Strategy Algorithm. In this paper, we summarize the previous research results [36], and experimentally compare and analyze the congested road sections of real data and one-day simulation data, among which there are 4 sections with the most apparent congestion characteristics. We select 4 high load segments for comparison, and the comparison is shown in Table 4.

From Tables 4 and it can be seen that the real-time data and simulation data show that the congestion period of the congested road section is consistent. In the congested road section, the flow rate, the average speed and the traffic density of the section are not different, in which Mean Absolute Error of the average speed is 1.075, Mean Absolute Error of the flow rate is 18.750, and Mean Absolute Error of the traffic density is 4.607. The experimental results are able to show the advantages of the simulation system in the

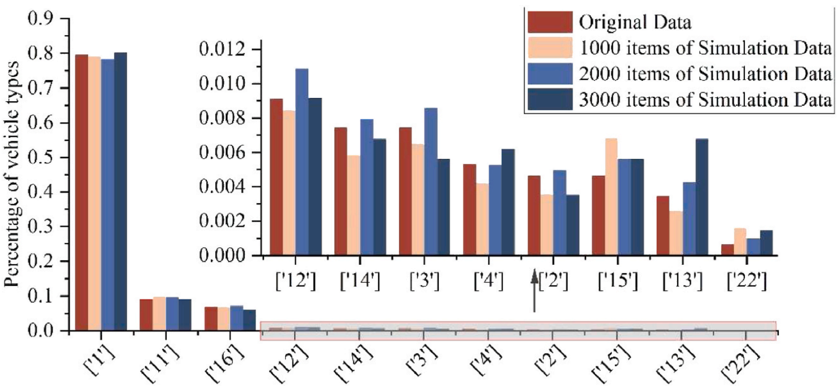

Fig. 12. Comparison of vehicle type distributions of different orders of magnitude of simulated data with real data.

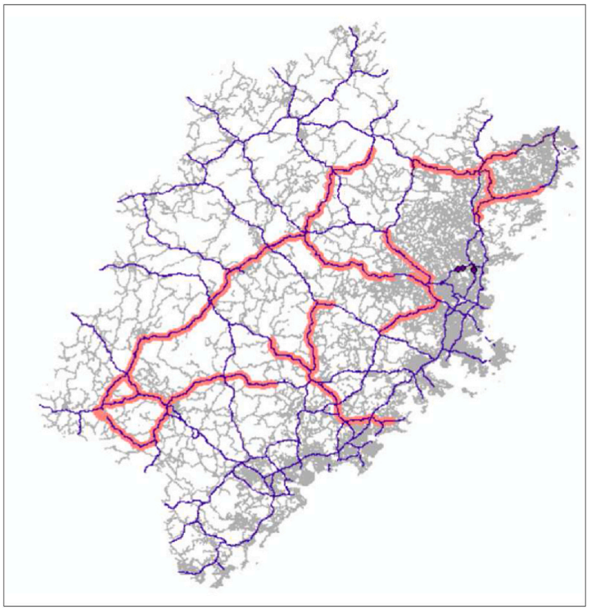

Fig. 13. Identification results of high-load road sections in a province at a certain time period.

Table 4
Comparison between simulated data and real data for some congested road sections.

Section	Data type	Congestion time period	Average speed	Flow	Traffic density
Section a	Actual	10:15:00–10:30:00	16.15	387	135.29
	Simulated	10:15:00–10:30:00	17.01	358	125.15
Section b	Actual	10:15:00–10:30:00	50.57	406	111.91
	Simulated	10:15:00–10:30:00	50.29	418	115.21
Section c	Actual	11:00:00–11:15:00	55.69	453	51.69
	Simulated	11:00:00–11:15:00	53.21	462	53.85
Section d	Actual	10:45:00–11:00:00	26.57	187	16.29
	Simulated	10:45:00–11:00:00	25.89	212	19.08

traffic prediction.

In order to simulate the situation of high load on the expressway network caused by concurrent traffic flow data, this experiment randomly generates 1–5 times the transaction flow through the simulation data, and uses Traffic Flow Control Strategy Algorithm to identify the high-load road sections that require optimal flow control. Fig. 14 shows the traffic flow distribution of these high-load road sections.

Fig. 14 shows the high-load road sections identified by the system. For each high-load road section, the system determines the toll stations and service areas that need to be coordinated and optimized for control, based on the road network topology map connection structure and considering the traffic flow at the expressway diversion convergence point. The system generates the current road network flow control strategy with a time step of 15 min to the manager. The manager can control the number of vehicles driving into and out of the road network according to the real-time traffic flow to optimize the overall traffic flow. These results provide valuable references for traffic management and planning of intelligent expressways, and are of great significance for improving traffic operation efficiency and service quality. It is worth noting that the optimization results obtained in this study are only applicable to a specific provincial area, and the strategy should be applied cautiously for other areas and verified and analyzed with local traffic data.

4.4. Multi-task scheduling analysis for traffic flow simulation

In order to effectively validate the performance of ETC_MTS in the simulation platform, we conducted experiments on three servers. Each server has eight computing nodes, each with 16 CPU cores and 32 threads. In our experiments, we set up task scenarios with 10, 50, 100, 500, 1000, and 2000 groups, where each group of tasks contains 3 parallel subtasks, by comparing with SJFS [37], Priority [38], FCFS [39], and RR [40], we evaluate their performance. Fig. 15 illustrates the comparison of the time efficiency of each algorithm.

The comparison of processing time and waiting time of each algorithm in multigroup tasks such as 10 groups, 50 groups, 100 groups, 500 groups, 1000 groups and 2000 groups is shown in Fig. 15. It is clear from the figure that ETC_MTS proposed in this paper

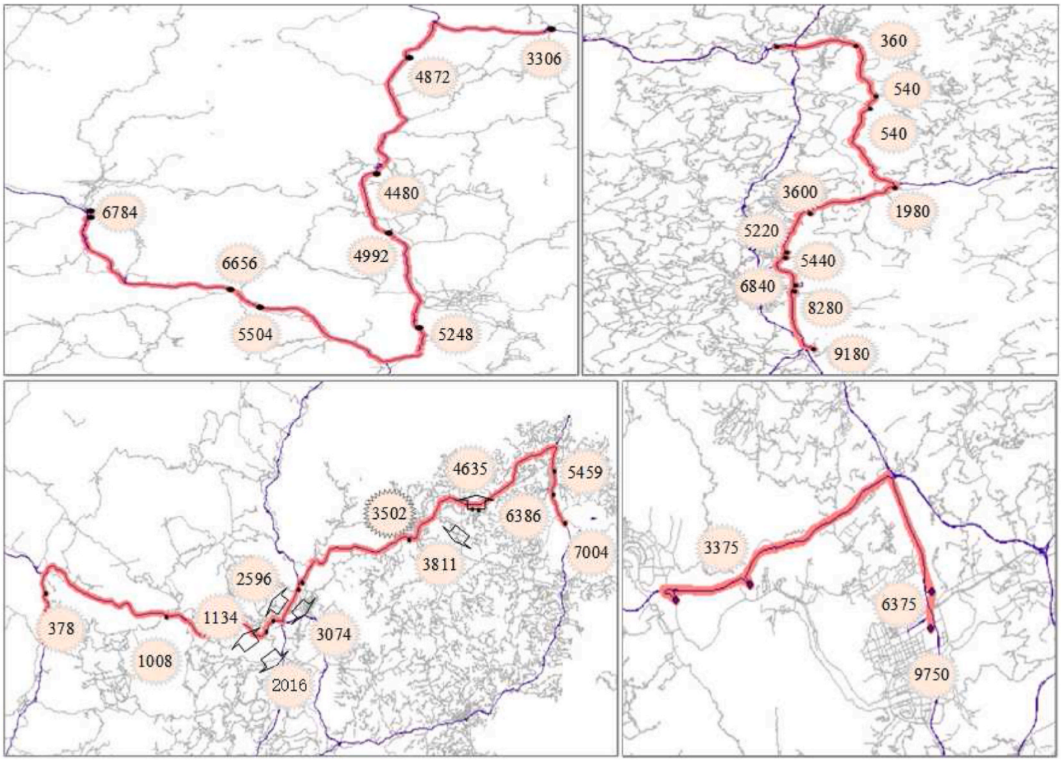

Fig. 14. Traffic flow distribution on different high loaded road sections.

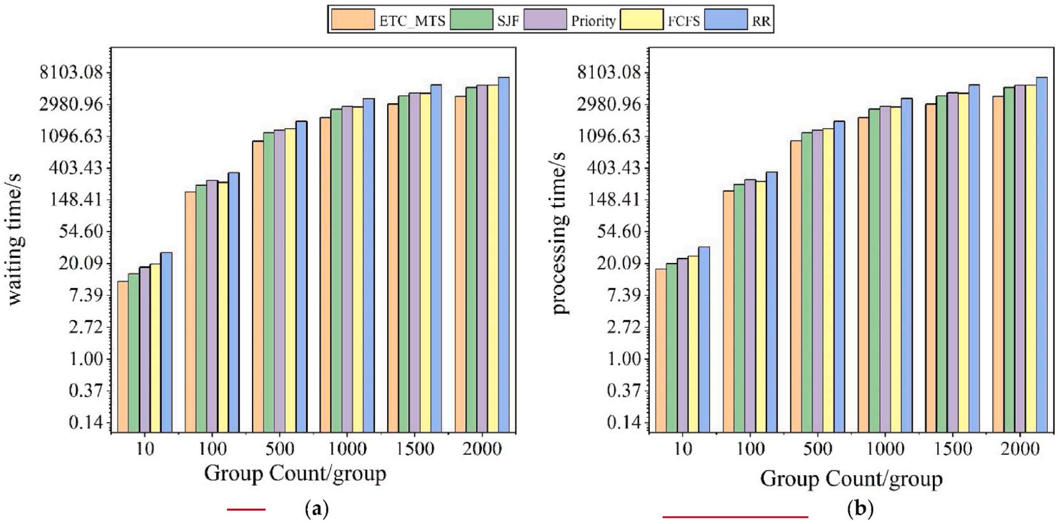

Fig. 15. (**a**) Comparison of waiting time for each algorithm to process different groups; (**b**) Comparison of processing time for each algorithm to process different groups.

outperforms the other algorithms in terms of processing time and waiting time.

Next, we show the algorithm's advantages more visually by calculating the time efficiency between each algorithm, which is shown in Fig. 16.

In Fig. 16, we can observe that with the increase of data volume, the combined processing time of multi-tasking algorithms such as SJFS, Priority, FCFS, and RR increases exponentially, while ETC_MTS shows higher time efficiency in comparison. Through the reasonable allocation and coordination of tasks, ETC_MTS can effectively reduce the execution time of tasks and the waiting time of multi-tasks, thus improving the efficiency of the overall simulation platform. These results verify the feasibility and practicality of

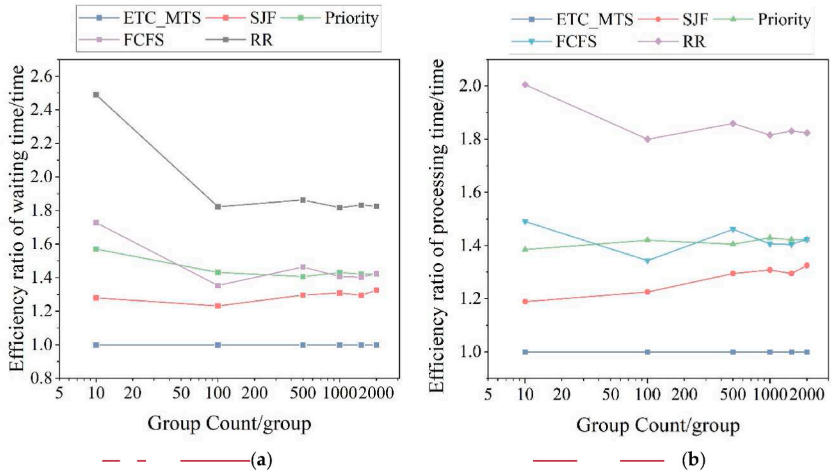

Fig. 16. (**a**) Comparison of efficiency ratio of waiting time; (**b**) Comparison of efficiency ratio of processing time.

ETC_MTS in traffic simulation platforms on expressways. However, it should be noted that different data volumes and task combinations may affect the performance of the algorithm, so it needs to be adjusted and optimized according to the specific situation in practical applications.

When executing concurrent subtasks, ETC_MTS algorithm sets the initial priority of each subtask to 0. Since subtasks need to compete for computational resources, the system evaluates the subtasks in a concurrent task based on their processing time, thread utilization, and waiting time of subsequent tasks, and then upgrades the priority of the subtasks in the concurrent task, and allocates computational resources to the higher priority subtasks when processing. In order to demonstrate the execution of ETC_MTS algorithm in the process of task scheduling processing in terms of processing time, waiting time and priority adjustment, the experiment is designed with 4 groups of concurrent tasks, each concurrent task needs to process 3 independent computational tasks in turn, and the computational tasks include calculating the in-transit, flow, saturation, commission, and average speed of vehicles and road conditions within the Block. Their operation is shown in Fig. 17.

In concurrent tasks, the system processes subtasks according to priority. According to the comprehensive evaluation results, the algorithm dynamically improves the execution priority of subtasks, so that subtasks can obtain resources first and complete tasks quickly. As shown in Fig. 17, when subtask 1 of concurrent task 1 is completed by 34 %, after the comprehensive evaluation of the system, computing resources are allocated to subtask 2 to make the system dynamically upgrade subtask 2. At this time, subtask 2 has higher priority than other subtasks, thus obtaining the execution privilege. When the subtask 2 is completed by 32 %, the system raises

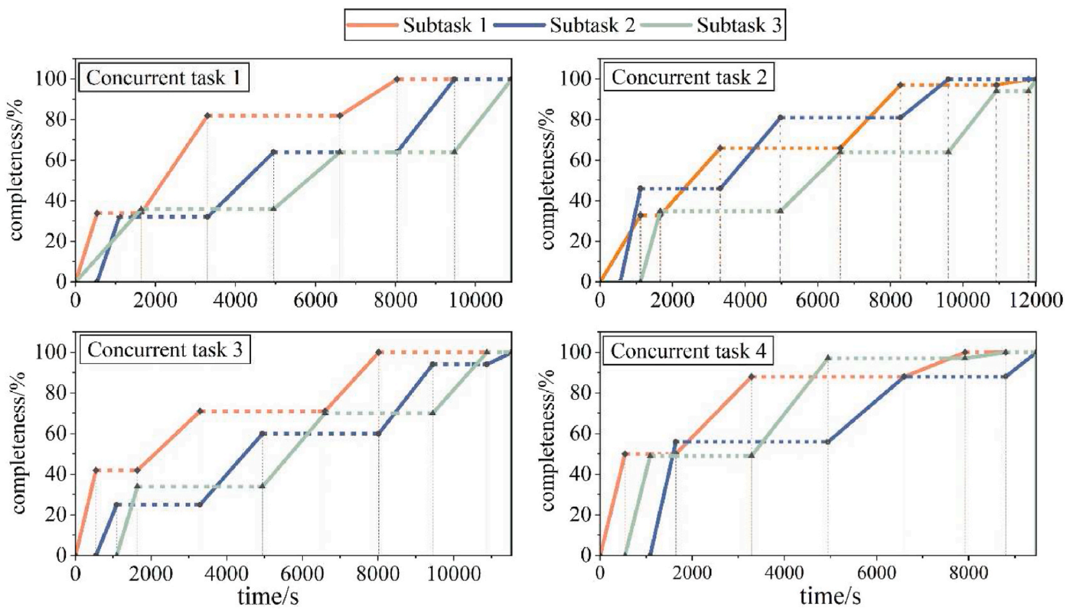

Fig. 17. Analysis of implementation of different concurrent task.

the priority of subtask 3, so that subtask 3 starts to be executed. Through continuous evaluation, the system preferentially allocates computing resources to subtasks with high execution efficiency until all three subtasks are completed.

5. Conclusions

In this paper, we propose a framework of multi-task real-time parallel simulation of ETC for intelligent expressway, which not only meets the demand of intelligent expressway simulation and verification system for ETC big data, but also realizes the quantitative calculation and display of real-time traffic and improves the intelligent management level of expressway. The following three key technologies are summarized in detail: 1) Considering multi-dimensional features such as vehicle type characteristics, traffic flow characteristics, toll station area location characteristics, time characteristics, and speed characteristics, we establish Feature Extraction Algorithm to realize the extraction of features in different dimensions of Edata., which can provide a basis for simulating Edata generation in the simulation and verification system. 2) Based on the real-time coordination of all kinds of traffic flow information generated by expressway in a certain period of time, and the demand of road network flow control, Traffic Control Strategy Algorithm is proposed, which can effectively simulate the situation of mass vehicles in the expressway network and provide traffic control recommendations for optimal control of real-time traffic flow. 3) Based on the concurrent task requirements of intelligent expressway simulation and verification system, ETC_MTS is established. Compared with the current mainstream multi-task allocation algorithms such as SJFS, Priority, FCFS and Round Robin, the execution performance of ETC_MTS in task analysis has been improved.

However, there is also room for improvement in this study. We will further use machine learning algorithms to correct the parameters of the feature extraction algorithm in our future research work, so that the Edata simulated by the simulation system will be more in line with the laws of society and continuously improve the reliability of the data in the simulation. Our next plan is to delve deeper into the area of expressway traffic flow scheduling, with a particular focus on the collection and analysis of data, such as the capacity of service areas and the number of exit/entrance lanes. This will provide more specific guidance for traffic scheduling and help optimize strategy development. At the same time, we will continue to improve the ETC simulation application system, broaden the ETC application scenarios, optimize the functions of the simulation system, and expand the application scope of the ETC simulation and verification system.

Funding

This work is funded by the relevant scientific research projects: the Renewable Energy Technology Research institution of Fujan University of Technology Ningde, China (Funding number: KY310338), the 2020 Fujian “One Belt, One Road” Science and Technology Innovation Platform Project (2020D002), Provincial Candidate Project of Fujian Province Million Talents Project (GY-Z19113), Patent Grant Project ((Funding number: GY-Z18081, GY-Z19099, GY-Z20074), Crosswise project (GY-H-20077), Municipal level science and technology projects (Funding number: GY-Z-22006, GY-Z-220230), Fujian Provincial Department of Science and Technology Foreign Cooperation Project (Funding num-ber:2023I0024), Research Platform Open Project (Funding number: KF-X19002, KF-19-22,001).

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Data availability statement

The authors do not have permission to share data.

CRediT authorship contribution statement

Fumin Zou: Validation, Resources, Funding acquisition, Formal analysis, Conceptualization. **Nan Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Feng Guo:** Software, Project administration, Investigation, Data curation. **Qiqin Cai:** Software. **Xinjian Cai:** Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Sun, Y. Huang, Y. Zhang, Thoughts on the development trend and construction of global smart expressway in the digital age, Highways (4) (2022) 237–242.
[2] F. Zou, F. Guo, J. Tian, S. Luo, X. Yu, Q. Gu, L. Liao, The method of dynamic identification of the maximum speed limit of expressway based on electronic toll collection data, Sci. Program. 2021 (2021) 1–15, https://doi.org/10.1155/2021/4702669.
[3] Z. Chen, F. Zou, F. Guo, Q. Gu, Short-term traffic flow prediction of expressway based on Seq2seq model, in: International Conference on Frontiers of Electronics, Information and Computation Technologies, 2021, pp. 1–5, https://doi.org/10.1145/3474198.3478239.
[4] J. Tian, F. Zou, F. Guo, Q. Gu, Q. Ren, G. Xu, Expressway traffic flow forecasting based on SF-RF model via ETC data, in: International Conference on Frontiers of Electronics, Information and Computation Technologies, 2021, pp. 1–7, https://doi.org/10.1145/3474198.3478238.
[5] F. Zou, Q. Ren, J. Tian, F. Guo, S. Huang, L. Liao, J. Wu, Expressway speed prediction based on electronic toll collection data, Electronics 11 (10) (2022) 1613, https://doi.org/10.3390/electronics11101613.
[6] S. Luo, F. Zou, C. Zhang, J. Tian, F. Guo, L. Liao, Multi-view travel time prediction based on electronic toll collection data, Entropy 24 (8) (2022) 1050, https://doi.org/10.3390/e24081050.
[7] Q. Cai, D. Yi, F. Zou, Z. Zhou, N. Li, F. Guo, Recognition of vehicles entering expressway service areas and estimation of dwell time using ETC data, Entropy 24 (9) (2022) 1208, https://doi.org/10.3390/e24091208.
[8] H. Chen, F. Zou, F. Guo, Q. Gu, X. Yu, Y. Luo, J. Xu, A ETC gantry information calibration method based on trajectory data of special transportation vehicles, in: International Conference on Frontiers of Electronics, Information and Computation Technologies, 2021, pp. 1–7, https://doi.org/10.1145/3474198.3478242.
[9] Y. Luo, F. Zou, F. Guo, Q. Gu, Z. Lin, Y. Lin, Spatial information extraction algorithm of ETC gantry based on trajectory mileage, in: International Conference on Frontiers of Electronics, Information and Computation Technologies, 2021, pp. 1–8, https://doi.org/10.1145/3474198.3478241.
[10] J. Wu, F. Zou, F. Guo, Q. Gu, S. Huang, Y. Luo, Research on detection of outlier point of highway ETC gantry based on SegrDTW mode, in: International Conference on Frontiers of Electronics, Information and Computation Technologies, 2021, pp. 1–8, https://doi.org/10.1145/3474198.3478240.
[11] F. Guo, F. Zou, S. Luo, H. Chen, X. Yu, C. Zhang, L. Liao, Positioning method of expressway ETC gantry by multi-source traffic data, IET Intell. Transp. Syst. (2022), https://doi.org/10.1049/itr2.12280.
[12] F. Zou, F. Guo, S. Luo, L. Liao, N. Li, Y. Xing, Research and design of expressway ETC simulation platform, J. Syst. Simul. (2022) 1–17.
[13] T. Alghamdi, S. Mostafi, G. Abdelkader, K. Elgazzar, A comparative study on traffic modeling techniques for predicting and simulating traffic behavior, Future Internet 14 (10) (2022) 294, https://doi.org/10.3390/fi14100294.
[14] G. Chun, N. Rouphail, M.S. Samandar, G. List, G. Yang, R. Akcelik, Analytical and microsimulation model calibration and validation: application to roundabouts under sight-restricted conditions, Transport. Res. Rec. 2677 (3) (2023) 274–288, https://doi.org/10.1177/03611981221115071.
[15] I.M. Septyaningrum, R.Y. Anindita, Traffic signalizing application at unsignalized intersection applying Vissim software microsimulation, RSF Conference Series: Eng. Technol. 2 (2) (2022) 294–306, https://doi.org/10.31098/cset.v2i2.583.
[16] H. Qin, W. Zhang, H. Zhai, Cooperative control of multiple intersections combining agent and chaotic particle swarm optimization, Comput. Electr. Eng. 110 (2023), 108875, https://doi.org/10.1016/j.compeleceng.2023.108875.
[17] F. Forouzandeh, H. Arman, A. Hadi-Vencheh, A.M. Rahimi, A combination of DEA and AIMSUN to manage big data when evaluating the performance of bus lines, Inf. Sci. 618 (2022) 72–86, https://doi.org/10.1016/j.ins.2022.10.044.
[18] Z. Cheng, L. Zhang, Y. Zhang, S. Wang, W. Huang, A systematic approach for evaluating spatiotemporal characteristics of traffic violations and crashes at road intersections: an empirical study, Transportmetrica: Transport. Sci. (2022) 1–23, https://doi.org/10.1080/23249935.2022.2060368.
[19] S. Khaleghian, H. Neema, M. Sartipi, T. Tran, R. Sen, A. Dubey, Calibrating real-world city traffic simulation model using vehicle speed data, in: 2023 IEEE International Conference on Smart Computing (SMARTCOMP), 2023, pp. 303–308, https://doi.org/10.1109/SMARTCOMP58114.2023.00076.
[20] M. Burger, M. Van Den Berg, A. Hegyi, B. De Schutter, J. Hellendoorn, Considerations for model-based traffic control, Transport. Res. C Emerg. Technol. 35 (2013) 1–19, https://doi.org/10.1016/j.trc.2013.05.011.
[21] Q. Chao, H. Bi, W. Li, T. Mao, Z. Wang, M.C. Lin, Z. Deng, A survey on visual traffic simulation: models, evaluations, and applications in autonomous driving, Comput. Graph. Forum 39 (1) (2020) 287–308, https://doi.org/10.1111/cgf.13803.
[22] W. Burghout, H.N. Koutsopoulos, Hybrid traffic simulation models: vehicle loading at meso-micro boundaries, in: Transport Simulation, EPFL Press, 2019, pp. 27–41.
[23] W. Shangguan, X. Li, L. Chai, Y. Cao, J. Chen, H. Pang, T. Rui, Research review on simulation and test of mixed traffic swarm in vehicle-infrastructure environment, J. Traffic Transport. Eng. 22 (3) (2022) 19–40, https://doi.org/10.19818/j.cnki.1671-1637.2022.03.002.
[24] J.W.C. Van Lint, S.C. Calvert, A generic multi-level framework for microscopic traffic simulation—theory and an example case in modelling driver distraction, Transp. Res. Part B Methodol. 117 (2018) 63–86, https://doi.org/10.1016/j.trb.2018.08.009.
[25] L. Huang, Research on Construction and Simulation of Typical Application Scenarios Based on Vehicle-Road Collaboration (Master's Thesis, Chongqing University, 2022, https://doi.org/10.27670/d.cnki.gcqdu.2021.001581.
[26] N. Zheng, W. Yang, L. Ma, H. Han, VISSIM simulation-based analysis and prediction of the effect of emergent traffic incidents on traffic operation of mountain highway, Saf. Environ. Eng. 27 (4) (2020) 223–230, https://doi.org/10.13578/j.cnki.issn.1671-1556.2020.04.030.
[27] F.A. Mullakkal-Babu, M. Wang, B. van Arem, B. Shyrokau, R. Happee, A hybrid submicroscopic-microscopic traffic flow simulation framework, IEEE Trans. Intell. Transport. Syst. 22 (6) (2020) 3430–3443, https://doi.org/10.1109/TITS.2020.2990376.
[28] V. Punzo, M. Montanino, A two-level probabilistic approach for validation of stochastic traffic simulations: impact of drivers' heterogeneity models, Transport. Res. C Emerg. Technol. 121 (2020), 102843, https://doi.org/10.1016/j.trc.2020.102843.
[29] M. Zhu, X. Wang, A. Tarko, Modeling car-following behavior on urban expressways in Shanghai: A naturalistic driving study, Transport. Res. C Emerg. Technol. 93 (2018) 425–445, https://doi.org/10.1016/j.trc.2018.06.009.
[30] Y. Yin, J. Zhao, Parameter calibration method of traffic flow model for traffic trajectory dispersion at intersections, J. Railw. Sci. Eng. (9) (2022) 2563–2573, https://doi.org/10.19713/j.cnki.43-1423/u.T20211073.
[31] B. Ciuffo, C.L. Azevedo, A sensitivity-analysis-based approach for the calibration of traffic simulation models, IEEE Trans. Intell. Transport. Syst. 15 (3) (2014) 1298–1309, https://doi.org/10.1109/TITS.2014.2302674.
[32] H. Wang, Design and Implementation of VISSIM Simulation Parameter Calibration System for Urban Road Traffic Based on Orthogonal Test Method, (Master's Thesis, Shijiazhuang Tiedao University), 2021, https://doi.org/10.27334/d.cnki.gstdy.2020.000319.
[33] J. Zhang, X. Wang, Research on traffic simulation model correction based on parameter sensitivity analysis, Appl. Res. Comput. (6) (2022) 1790–1795, https://doi.org/10.19734/j.issn.1001-3695.2021.11.0593.
[34] Kai Ma, Research on Private Car Trajectory Data Generation Based on Spatio-Temporal Interaction (Master's Thesis, Dalian University of Technology, 2021, https://doi.org/10.26991/d.cnki.gdllu.2020.001171.
[35] X. Kong, Q. Chen, M. Hou, A. Rahim, K. Ma, F. Xia, RMGen: a tri-layer vehicular trajectory data generation model exploring urban region division and mobility pattern, IEEE Trans. Veh. Technol. 71 (9) (2022) 9225–9238, https://doi.org/10.1109/TVT.2022.3176243.
[36] M.Z. Mehdi, H.M. Kammoun, N.G. Benayed, D. Sellami, A.D. Masmoudi, Entropy-based traffic flow labeling for CNN-based traffic congestion prediction from meta-parameters, IEEE Access 10 (2022) 16123–16133, https://doi.org/10.1109/ACCESS.2022.3149059.
[37] R.K. Mondal, E. Nandi, D. Sarddar, Load balancing scheduling with shortest load first, Int. J. Grid and Distribut. Comput. 8 (4) (2015) 171–178, https://doi.org/10.14257/ijgdc.2015.8.4.17.
[38] S. Ghanbari, M. Othman, A priority based job scheduling algorithm in cloud computing, Procedia Eng. 50 (2012) 778–785.
[39] N. Bansal, A. Maurya, T. Kumar, M. Singh, S. Bansal, Cost performance of QoS Driven task scheduling in cloud computing, Procedia Comput. Sci. 57 (2015) 126–130, https://doi.org/10.1016/j.procs.2015.07.384.
[40] P. Pradhan, P.K. Behera, B.N.B. Ray, Modified round robin algorithm for resource allocation in cloud computing, Procedia Comput. Sci. 85 (2016) 878–890, https://doi.org/10.1016/j.procs.2016.05.278.