International Journal of Frontiers in Engineering Technology
ISSN 2706-655X Vol.4, Issue 4: 83-88, DOI: 10.25236/IJFET.2022.040411

Distributed Intelligent Transportation System Data Management and Processing Platform

Xiaobing Peng*

North China University of Technology, Beijing, China
p18090472854@163.com
**Corresponding author*

Abstract: *In recent years, with the rapid development of China's road construction and the continuous improvement of people's living standards, the number of urban motor vehicles has continued to grow year after year. Motor vehicles have become one of the main means of transportation for urban residents. At the same time, due to the increasing amount of traffic information data, a single server platform can not meet the reception and storage of massive traffic data, nor can it meet the needs of rapid query of massive traffic data. Therefore, this paper will use the distributed system to process data to solve the above problems, so as to ensure the real-time transmission and storage of massive traffic data. This paper discusses and analyzes the characteristics of distributed intelligent transportation system and the design of communication system data management and processing platform. Through the behavior decision algorithm of distributed transportation system, the flow data of distributed intelligent transportation system are tested and analyzed. The experimental results show that in the time period from 6:00 to 8:00, from 12:00 to 14:00 and from 16:00 to 18:00 when the traffic data reaches the peak, the amount of data processed by each communication server is very different, and the difference data amount can reach 60000 traffic data. The statistical results of this traffic data amount can show that the ratio of the data amount sent by each front-end bayonet every moment is not equal to the ratio of the data amount of each front-end bayonet every day, which is enough to show that the data management and processing capacity and efficiency of the distributed intelligent transportation system have been greatly improved.*

Keywords: *Distributed System, Intelligent Transportation System, Data Management Design, Processing Platform Design*

1. Introduction

With the increasingly prominent traffic contradictions, road construction alone can not fundamentally solve the problem. Therefore, the intelligent transportation based on big data technology, cloud technology and IOT technology has been recognized and concerned by the society. The core of intelligent traffic management system is highway operation management, which is based on computer network and information management and communication system. It is an important support system for highway and traffic infrastructure. In short, intelligent traffic management system plays a very important role in traffic and transportation safety.

Many scholars at home and abroad have studied the design of data management and processing platform of distributed intelligent transportation system. Manias DM proposed a collaborative distributed intelligent technology - federated learning. Through its case study, it emphasizes the ability of the federated model deployed on the roadside infrastructure of the whole network to recover from failure by using group intelligence, while reducing recovery time and restoring acceptable system performance. Federal learning has many use cases and benefits. It is a key driving factor of intelligent transportation system and is expected to be widely implemented in 5g and other networks and applications [1]. Mukherjee a proposed a multi input multi output (MIMO) technology model in wireless sensor networks. The model uses back propagation neural network (BPNN) to solve the problem of cluster head recognition in MIMO sensor networks. Due to the dynamic and real-time nature of the environment, the traditional channel recognition lacks location recognition. Therefore, in order to obtain more accurate positioning accuracy, BPNN combined with distributed gradient descent technology is used to calculate the position of unknown ch. This reduces the distance estimation error, and further uses particle swarm optimization technology to obtain the optimal weight and threshold of the network. This work is verified by mathematical analysis, simulation and comparison with the existing technology. The model shows better performance in terms of energy consumption, error rate and calculation time [2].

Intelligent transportation system integrates communication and transportation technology, covering a wide range of contents. Intelligent transportation is a real-time and forward-looking intelligent traffic management method based on the Internet of things technology, which collects massive data through front-end equipment, counts, summarizes and analyzes these data, and then gives suggestions and guidance according to the analysis results. In this way, how to effectively store and process has become a difficult task. Therefore, a computing technology that can carry the storage and processing of big data has become a key factor. By using the distributed architecture of the cloud, a huge amount of data can be distributed to all nodes in the cloud, which greatly improves both the reading and writing speed and processing performance, and will continue to improve with the continuous expansion of the scale of the cloud [3, 4].

2. Data Management and Processing of Distributed Intelligent Transportation System

2.1. Characteristics of Distributed System

Distributed system is a system whose hardware or software components are distributed on networked computers, and the components communicate and coordinate their actions by passing messages. Distributed system is composed of networked computers, and computer cluster is not only the basis of constructing distributed system, but also an important part. In the distributed system, each node is transmitted through messages. It is a system in which components complete communication and action coordination through message transmission on networked computers. Therefore, distributed system has three characteristics: concurrency of components, lack of global clock and independence of component failure. Specifically, the three basic characteristics of distributed system are concurrency, scalability and fault independence [5, 6].

2.1.1. Concurrency

Concurrency refers to the concurrent execution of each thread in the program, such as concurrent transmission or concurrent processing of messages.

2.1.2. Flexibility

The distributed system can still ensure the efficient processing performance of the system itself under the condition of increasing the level of data volume, which is the advantage of the distributed system [7].

2.1.3. Fault Independence

Compared with the traditional C / S structure, the distributed system can not only send massive data from the client and the server can quickly return the results to the client [8, 9], but also avoid the failure of system nodes due to some external or internal reasons. In order to avoid the data processing failure caused by these faults, the distributed system needs to have the characteristics of fault independence.

2.2. Transportation System Data Management and Processing Platform

Real time: the system is to process traffic data in real time. First, ensure that the system processes the data packets sent by the client in real time, and make the application server query vehicle information and see vehicle data and vehicle pictures in real time. The so-called real-time system is a system that can respond to and process the generated task events within a limited time interval. In this system, the real-time processing of task generation trigger thread is mainly used. The processing time of each task in this system is limited and can not affect the processing of subsequent data. The so-called processing needs a good real-time scheduling algorithm to ensure the time limit and high reliability of the system at the time of confirmation. After each traffic data is received, the system needs to analyze the data packet, then store it in the message buffer queue inside the system, and then trigger a new forwarding thread to start according to the data in the message queue to realize the forwarding task between the nodes inside the system [10, 11].

Fault tolerance: the real-time processing system has high requirements for system fault tolerance, so the most important point of the distributed system designed in this paper is to ensure that it can still

ISSN 2706-655X Vol.4, Issue 4: 83-88, DOI: 10.25236/IJFET.2022.040411

process real-time traffic data normally in the case of data error or system node error. The errors faced by the distributed system are mainly reflected in the two aspects of data security and node failure. For the problem of data packet error, it is necessary to analyze the data packet, and then inform the front end of the error information and require data retransmission. For the problem of node failure and error, it is necessary to forward the packet message to the effective processing node to avoid the loss of data caused by blocking [12].

3. Behavior Decision Algorithm of Distributed Transportation System

It is an important problem in the decision-making process of car following behavior to choose appropriate rules to make a decision on the speed of car following at t + 1 time. Because different attributes have different influence on the selection of decision rules, the weight of each attribute in the decision rule table needs to be considered when following the vehicle speed decision. In the decision-making process of car following behavior, the speed of the following car at t + 1 time is determined according to the mean value of the following car speed in the selected rule. According to the observation value p (T) of the following car at time t, we calculate the weighted distance from the observation point to each rule in the rule table, and select the decision rule whose weighted distance is less than the given threshold to calculate the speed of the following car at time t + 1. The calculation method of weighted distance is shown in formula (1):

$$dist(p(t),r)=\sqrt{\sum_{s\in S_{opt}} w_s(pv(s,t)-rv(s))^2} \tag{1}$$

Where s is the attribute in the conditional attribute set CoPt, PV (s, t) represents the value of attribute C in the observation point P (T), RV (s) is the value of attribute s in rule R, and WC is the weight of attribute C in the decision rule table. The weight can be expressed by the discrimination of attribute C in the condition attribute set. For attributes $\mathrm{c}\in C_{opt}$, their discrimination is defined in the form of formula (2):

$$\mathrm{K}_{\mathrm{S_{ept}}}(s)=\frac{\left|H/S_{opt}\right|-\left|H/(S_{opt}-\{S\})\right|}{\left|H/S_{opt}\right|} \tag{2}$$

Where $K/C_{\mathrm{opt}}=K/ind(C_{opt})=\{X_1,X_2,...,X_{\mathrm{n}}\}$, $\left|K/C_{opt}\right|$ represents the number of equivalent classes on the attribute set CoPt, and the weight of attribute C is calculated according to formula (3):

$$w(s)=K_{C_{opt}}(s)/\sum_{s\in s_{opt}} p_{S_{opt}(s)} \tag{3}$$

4. Test and Analysis of Distributed Intelligent Transportation System Data

In the experimental environment, we set the front-end bayonet client to send 3 million traffic data every day to simulate the data flow in the real environment. Do an experiment according to the configuration file updated once a day. Firstly, the simulation front-end bayonet client and distributed system server are built; On the first day of the experiment, when the system is started, the system configuration file is not initialized, the data is forwarded in a random way and forwarded to the corresponding communication server, and the data processing of each communication server on the first day is obtained;. On the second day of the experiment, analyze the amount of data sent by each front-end bayonet in the database on the first day, write the front-end bayonet IP address and the corresponding communication server IP address into the configuration file, and then start the system reading the configuration file thread to initialize the configuration file. The system message scheduling model determines which communication server the real-time data should be forwarded to for processing according to the corresponding relationship in the configuration file, Thus, the data processing situation of each communication server on the next day is obtained; After sending the data on the day of the experiment, the process of updating the configuration file starts again to analyze the data in the database and update the configuration file; Finally, repeat steps 3 and 4. Get the data flow of

ISSN 2706-655X Vol.4, Issue 4: 83-88, DOI: 10.25236/IJFET.2022.040411

each communication server under this model. The comparison is shown in Figure 1.

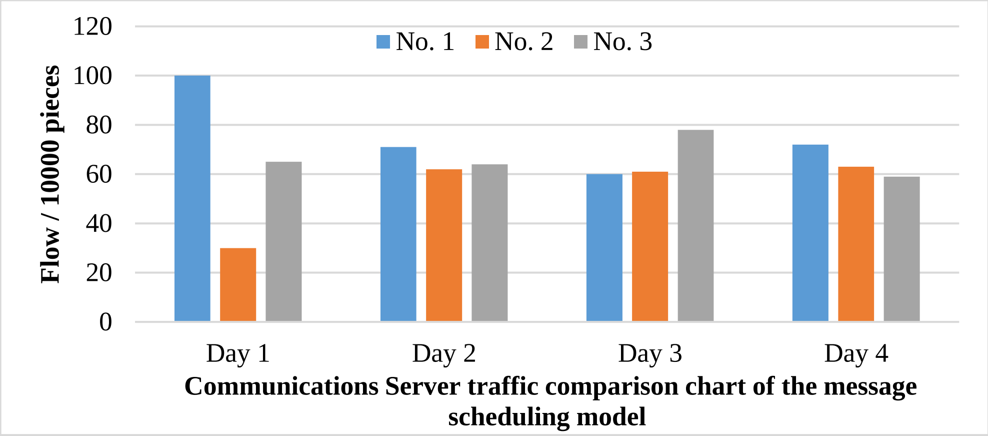

Figure 1: Communications Server traffic comparison chart of the message scheduling model

It can be seen from the data in the figure above that when the configuration file under this model is not used, the total amount of data processed by each communication server is different every day, and the difference in the amount of data is relatively large. As a result, the pressure of communication servers with large amount of data is significantly higher than that of other communication servers. On the contrary, when the configuration file under this model is added, the amount of data processed by each communication server is roughly the same every day. The experimental results can show that the pressure of the amount of data processed by the communication server every day is roughly the same. Such experimental results are not close, which can not explain that the communication pressure of each communication server is roughly the same at all times. Therefore, we will measure the time of updating the configuration file and record the amount of data processed by each server at different time intervals during the peak traffic.

Before starting to verify whether the time interval of updating the configuration file affects the processing data of each communication server, first make a data statistics on the massive traffic data. Now take one hour as the time interval to count the amount of traffic data received and processed by each communication server in one hour. This experimental data is directly counted from the database based on the experimental data described in Figure 1. According to the statistics of the traffic data processed by each communication server per hour, the server described is the communication server, the abscissa is the statistical time period, the unit is hour, and the ordinate is the statistical traffic data, the unit is 10000, as shown in Table 1 and Figure 2.

Table 1: Sharing a single communications server traffic statistics

	2	4	6	8	10	12	14	16	18	20	22	24
Server 1	1.6	2.2	2.5	12.1	4.9	8.0	8.2	4.0	9.1	7.0	4.0	3.8
Server 2	1.7	1.9	2.0	9.2	3.8	12.0	12.9	3.7	6.9	6.0	5.1	2.0
Server 3	1.2	1.6	2.0	13.1	2.0	4.9	15.1	2.0	4.0	10.0	5.8	4.0

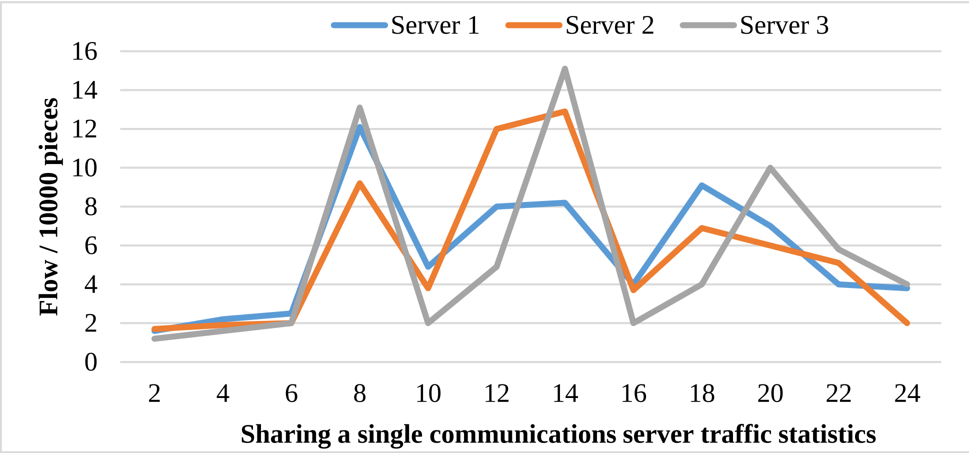

Figure 1: Sharing a single communications server traffic statistics

ISSN 2706-655X Vol.4, Issue 4: 83-88, DOI: 10.25236/IJFET.2022.040411

It can be seen from the chart that in the time period from 6:00 to 8:00, from 12:00 to 14:00 and from 16:00 to 18:00 when the traffic data volume reaches the peak, the data volume processed by each communication server is very different, and the difference data volume can reach 60000 traffic data. The statistical results of this traffic data volume can show that, The ratio of the amount of data sent by each front-end bayonet at every moment is not equal to the ratio of the amount of data of each front-end bayonet in a day. This paper changes the front-end computer configuration file in real time according to the actual situation, so as to make the data processing pressure of each server roughly equal every minute or even every second. Through the analysis and experimental verification of the model, it is verified that the profile update time formulated by the model under the condition of experimental verification. After the dynamic update of the profile, the data volume pressure processed by each communication server is roughly the same, so as to ensure the ability of real-time forwarding and processing of traffic data.

5. Conclusions

This paper studies the distributed data management and processing of urban transportation system, and has achieved some phased research results. There are still deficiencies in these research results, which need to be further discussed and improved in the follow-up research work. In the data-driven car following model based on rough set theory and vehicle trajectory reconstruction, it is necessary to improve the distributed traffic data management system to enable it to deal with the real-time processing of massive data; In the process of car following behavior decision and trajectory reconstruction decision, it is necessary to consider the impact of typical data on the decision algorithm, and further improve the rule extraction and decision algorithm design based on rough set, so as to improve the efficiency of the algorithm. In the aspect of distributed traffic optimal control based on back pressure algorithm, the problem of optimal data management in the case of asynchronous switching of intersection signals needs to be further considered; In the hierarchical multi granularity optimization control, a reasonable control sub area boundary control algorithm is designed to realize the coordination of internal control and boundary control in the control sub area, so as to improve the hierarchical multi granularity optimization control strategy of urban road traffic system.

References

[1] Manias D M., Shami A., Making a Case for Federated Learning in the Internet of Vehicles and Intelligent Transportation Systems. IEEE Network, 2021, 35(3):88-94.

[2] Mukherjee A., Jain D K., Goswami P., et al. Back Propagation Neural Network based Cluster Head Identification in MIMO Sensor Networks for Intelligent Transportation Systems. IEEE Access, 2020, 8(1):28524-28532.

[3] Lee Y., Jeong S., Masood A., et al. Trustful Resource Management for Service Allocation in Fog-Enabled Intelligent Transportation Systems. IEEE Access, 2020, (99):1-1.

[4] Balfaqih M., Ismail M., Nordin R., et al. Fast handover solution for network-based distributed mobility management in intelligent transportation systems. Telecommunication Systems, 2017, 64(2):1-22.

[5] Dabiri A., Kulcsar B., Distributed Ramp Metering—A Constrained Discharge Flow Maximization Approach. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(9):2525-2538.

[6] Rezaei M., Noori H., Razlighi M M., et al. ReFOCUS+: Multi-Layers Real-Time Intelligent Route Guidance System with Congestion Detection and Avoidance. IEEE Transactions on Intelligent Transportation Systems, 2019, (99):1-14.

[7] Ao L., Cruickshank H., Yue C., et al. Blockchain-Based Dynamic Key Management for Heterogeneous Intelligent Transportation Systems. IEEE Internet of Things Journal, 2017, (99):1-1.

[8] Usman M., Jan M A., Jolfaei A., SPEED: A Deep Learning Assisted Privacy-Preserved Framework for Intelligent Transportation Systems. IEEE Transactions on Intelligent Transportation Systems, 2020, (99):1-9.

[9] Gusrialdi A., Qu Z., Simaan M A., Distributed Scheduling and Cooperative Control for Charging of Electric Vehicles at Highway Service Stations. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(10):2713-2727.

[10] Ning Z., Sun S., Wang X., et al. Blockchain-enabled Intelligent Transportation Systems: A Distributed Crowdsensing Framework. IEEE Transactions on Mobile Computing, 2021, (99):1-1.

ISSN 2706-655X Vol.4, Issue 4: 83-88, DOI: 10.25236/IJFET.2022.040411

[11] Yoon S C., Shin T S., Lawrence K., et al. Development of Online Egg Grading Information Management System with Data Warehouse Technique. Applied Engineering in Agriculture, 2020, 36(4):589-604.
[12] Husain M., Y Alsaawy, Tufail A., A seven tier architecture of cloud database management system. Journal of Engineering and Applied Sciences, 2018, 13(13):5084-5089.