Theory and Practice of Science and Technology
2025, VOL. 6, NO. 4, 184-190
DOI: 10.47297/taposatWSP2633-456933.20250604

A Study On The Application Of Big Data Technology In Enhancing The Efficiency Of Intelligent Traffic Flow Management

Zhenxiang Guan, Hongyan Sun

Liaoning University of Science and Technology, Anshan, Liaoning 114051

ABSTRACT
This paper discusses the application of big data technology in intelligent traffic flow management, aiming at analyzing the framework of big data technology, the specific implementation process, the main algorithms and experimental evaluation, revealing its important role in improving the efficiency of traffic flow management. The study shows that big data technology can effectively integrate traffic data resources, optimize traffic signal control, realize accurate prediction and real-time regulation of traffic flow, thus significantly improving road capacity, easing traffic congestion, and providing powerful support for the development of intelligent transportation system.

KEYWORDS
Big data technology; Intelligent transportation; Traffic management; Efficiency enhancement; Algorithm application that

1 Introduction

This paper focuses on the application of big data technology in enhancing the efficiency of intelligent traffic flow management, aiming to discuss how big data technology can provide powerful support for traffic flow management, so as to realize accurate prediction, real-time regulation and optimization of traffic flow. Through the use of big data technology, we can integrate massive traffic data resources, explore the valuable information contained therein, and then provide a scientific basis for traffic signal control, traffic guidance information release, etc., to effectively alleviate traffic pressure, improve travel efficiency, and reduce the probability of traffic accidents [1].

The study of this topic is of great practical significance. From the social point of view, improving the efficiency of traffic flow management can significantly improve the urban traffic situation, reduce people's travel time and cost, improve the quality of life, and promote the sustainable development of the city. From the economic point of view, easing traffic congestion can reduce logistics and transportation costs, improve the operational efficiency of enterprises, and enhance the economic vitality and competitiveness of the city. From the environmental point of view, reducing unnecessary idling and frequent starting and stopping of vehicles can help reduce energy consumption and exhaust emissions, protect the ecological environment, and promote the development of green transportation. Therefore, this study has far-reaching significance for promoting the construction and development of intelligent transportation system and building an efficient, safe and environmentally friendly transportation system [2].

2 The main technological framework

2.1 Big data technology system

The big data technology system covers a number of key technical aspects such as data acquisition, storage, processing and analysis, which are described in detail below, as shown in Figure 1.

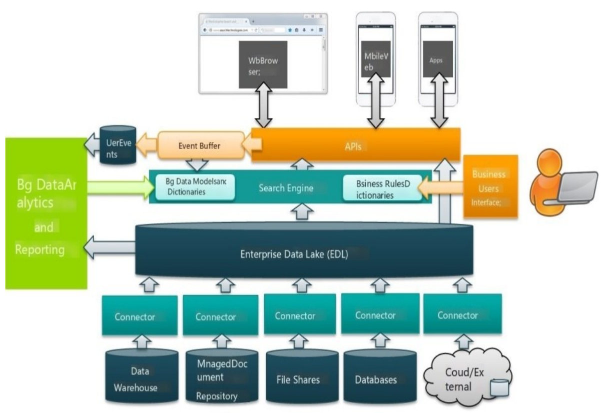

Figure 1 Big data architecture

Data acquisition

Hardware devices: Traffic cameras, geomagnetic sensors, microwave radar, GPS devices, etc. are the main hardware devices for traffic data collection. These devices are distributed in key locations such as city roads, highways, and traffic hubs to collect real-time information on vehicle speed, flow, and location [3].

Data interface: Through IoT technology, various hardware devices are connected to the data collection system to realize automatic data transmission. For example, using wireless communication technology (e. g., 4G/5G, Wi-Fi) to transmit video data collected by cameras to the data center.

Data preprocessing: During the acquisition process, data are initially processed, such as removing noisy data, filling in missing values, etc., to ensure the completeness and accuracy of the data [4-5].

Data storage

Distributed storage: Distributed file systems (such as Hadoop Distributed File System, HDFS) and distributed databases (such as Apache Cassandra) are used to store massive data. These systems are able to decentralize data storage in multiple nodes to improve storage capacity and data read and write speed.

Data warehouses and data lakes: data warehouses are used to store structured data to facilitate complex data analysis and query; data lakes can store various types of data, including unstructured data, and provide rich data resources for data mining and machine learning [6].

Cloud storage: the use of cloud computing platforms (such as Amazon S3, AliCloud OSS) to provide elastic storage services, according to the amount of data to dynamically adjust the storage resources to reduce storage costs.

Data processing

Batch processing: using MapReduce and other frameworks for offline processing of large-scale data, suitable for data analysis tasks that do not require high real-time performance. For example, statistical analysis of historical traffic flow data to generate traffic flow reports [7].

Stream processing: using Apache Kafka, Apache Flink and other stream processing frameworks, real-time data for rapid processing and analysis to achieve real-time monitoring of traffic flow and early warning. For example, when a section of traffic congestion, the system can issue real-time alerts and take appropriate measures.

Data cleansing and fusion: the collected data from multiple sources are cleaned to remove repetitive data, erroneous data, etc.; at the same time, different types of data are fused to form a unified data format for subsequent analysis.

Data analysis

Data Mining: Data mining algorithms such as cluster analysis, classification analysis and association rule mining are used to extract valuable information from massive data. For example, traffic flow data can be classified into peak hours and off-peak hours through cluster analysis, which provides the basis for traffic signal control [8-9].

Machine Learning and Deep Learning: Use machine learning algorithms (e.g., decision trees, support vector machines) and deep learning models (e. g., convolutional neural networks CNN, long and short-term memory networks LSTM) to model and predict traffic data. For example, the deep learning model predicts the traffic flow changes in the future period and adjusts the traffic signal timing in advance [10].

Visual analysis: the results of the analysis through visualization tools (such as Tableau, PowerBI) to display, intuitively present the trend of traffic flow, congestion area distribution and other information, to facilitate decision-making by traffic management personnel.

2.2 Intelligent Transportation System Architecture

Intelligent Transportation System (ITS) is a complex integrated system, and its overall architecture usually includes sensing layer, transmission layer, processing layer and application layer, and the following is a detailed description of each layer.

The perceptual layer

Function: Perception layer is the foundation of intelligent transportation system, responsible for real-time collection of traffic information, including vehicle location, speed, flow, road conditions. The main devices are traffic cameras, geomagnetic sensors, microwave radar, GPS devices, RFID tags and so on.

Equipment layout: traffic cameras are usually installed at intersections, highway entrances and exits, bridges and other key locations for the collection of vehicle images and video data; geomagnetic sensors and microwave radar are buried in the road below or on the side of the road, real-time monitoring of vehicular traffic and speed; GPS devices are installed in the vehicle to provide real-time location of the vehicle information.

Data Acquisition: These devices transmit the collected data to the transmission layer through the sensor network to provide basic data for subsequent data processing and analysis.

Transport Layer

Function: The transmission layer is responsible for transmitting the data collected in the sensing layer to the processing layer, and at the same time feeding back the instructions from the processing layer to the sensing layer devices. Transmission methods include wired transmission (e.g. fiber optic communication) and wireless transmission (e.g. 4G/5G, Wi-Fi, Bluetooth, etc.).

Network architecture: IoT technology is used to build a transmission network to ensure the stability and real-time data transmission. For example, the low-latency feature of 5G network is used to realize the real-time transmission of traffic camera images and the rapid issuance of traffic signal control commands.

Data security: In the transmission process, the data is encrypted to prevent data leakage and tampering and to ensure the security and integrity of the data.

Processing layer

Function: Processing layer is the core of ITS, responsible for storing, processing and analyzing the massive data collected. The main technologies include big data storage technology (e.g. HDFS, Cassandra), data processing framework (e.g. MapReduce, Flink) and data analysis algorithms (e.g. data mining, machine learning).

Data processing process: first of all, the collected data are cleaned and pre-processed to remove noise data and repetitive data; then different types of data are fused to form a unified data format; then the data analysis algorithms are used to mine and analyze the data to extract valuable information; finally, the analysis results are stored in the data warehouse or data lake for use by the application layer.

Decision-making support: According to the analysis results, generate traffic management decision-making instructions, such as traffic signal control scheme, traffic guidance information release, etc., and will be transmitted to the sensory layer equipment to achieve real-time regulation of traffic flow.

The application layer

Function: The application layer is the interface between ITS and users, responsible for providing the decision-making information generated by the processing layer to traffic managers, drivers and the public. The main applications include traffic signal control, traffic guidance information release, electronic toll collection, vehicle navigation and information services.

Traffic signal control: According to traffic flow data, dynamically adjust the timing parameters of traffic signals to optimize intersection capacity and reduce vehicle waiting time. For example, increase the green time of main roads during peak hours to ease traffic congestion.

Traffic guidance information release: through electronic display screens, cell phone APP, radio and other means, to provide drivers with real-time road information, the best path planning and congestion warning, to guide vehicles to avoid congested road sections, and improve the efficiency of road access.

Electronic toll collection: the use of RFID technology to realize the non-stop charging of vehicles, improve the efficiency of the toll station and reduce the queuing time of vehicles.

Vehicle navigation and information services: provide drivers with real-time navigation services, including road information, surrounding facilities inquiries (e.g., gas stations, parking lots, hospitals, etc.), to improve the convenience and safety of travel.

3 Specific realization processes

3.1 Data acquisition and pre-processing

3.1.1 Methods of collecting traffic flow data

Hardware equipment selection

Traffic camera: Traffic camera is one of the important equipment to collect traffic flow data. High-definition cameras can capture the real-time road vehicles traveling, through the image recognition technology, can accurately count the traffic flow, speed and vehicle type and other information. For example, in the city's major arteries and highways installed at the entrance and exit of the high-definition camera, real-time monitoring of the vehicle through the situation, to provide basic data for traffic flow analysis.

Geomagnetic sensors: Geomagnetic sensors sense the passage of vehicles by detecting their interference with the earth's magnetic field. It is usually buried under the road, when the vehicle passes by, will cause changes in the magnetic field, thus triggering the sensor to record the vehicle's passing time, speed and other information. Geomagnetic sensors have the advantages of easy installation and low cost, and are suitable for monitoring intersections and road sections with large traffic flow.

Microwave radar: Microwave radar utilizes the principle of microwave signal reflection to detect vehicles. It can measure the speed, distance and number of vehicles and other information, with high detection accuracy and anti-interference ability. Microwave radar can be installed on both sides of the road or the central isolation zone, real-time monitoring of traffic flow in multiple lanes, for traffic signal control and flow analysis to provide accurate data support.

GPS devices: GPS devices are installed on vehicles, which can obtain real-time information on the location, speed and driving direction of vehicles. By collecting GPS data from a large number of vehicles, we can analyze the distribution of traffic flow, the driving path of vehicles and congestion. For example, the GPS data of cabs and online taxis can provide rich data resources for the analysis of urban traffic flow, and help the traffic management department to understand the traffic situation in different areas.

Principles of layout

Coverage of key areas: In urban traffic flow monitoring, priority should be given to the layout of data collection equipment in areas with high traffic flow and easy congestion. For example, traffic cameras and geomagnetic sensors should be installed in key areas such as main roads in urban centers, transportation hubs (e.g., railway stations, airports), and the periphery of commercial centers, in order to comprehensively grasp the changes in traffic flow in these areas.

Reasonable distribution of monitoring points: According to the structure of the road network and the distribution characteristics of traffic flow, the location and number of monitoring points should be reasonably determined. The distance between the monitoring points should be determined according to the length of the road and the fluctuation of traffic flow, generally set up a monitoring point every 500 meters to 1,000 meters on the city's main roads to ensure that the trend of traffic flow can be accurately captured.

Consideration of data complementarity: different types of collection equipment have their own advantages and limitations, so the data complementarity between equipment should be considered in the layout. For example, the combination of traffic cameras and geomagnetic sensors can simultaneously obtain vehicle image information and speed, flow and other data to improve the accuracy and reliability of the data.

3.1.2 Key steps in data pre-processing

Data cleansing

Noise data removal: Traffic data acquisition process may be subject to various factors of interference, resulting in noise data. For example, traffic cameras may capture blurred images due to bad weather, low light or equipment failure, resulting in vehicle identification errors; geomagnetic sensors may be interfered by surrounding metal objects, resulting in false trigger signals. In order to ensure the quality of data, it is necessary to detect and remove noise from the collected data. Image processing algorithms can be used to denoise the images collected by the traffic camera to improve the clarity and recognizability of the images; for the data collected by the geomagnetic sensor, reasonable thresholds and filtering algorithms can be set to remove the abnormal signals generated by interference.

Filling in missing values: In the process of data acquisition, some data may be missing due to equipment failure, communication interruption and other reasons. The existence of missing values will affect the integrity of the data and the accuracy of the analysis results, so the missing values need to be filled. Filling methods can be selected according to the characteristics of the data and the missing situation, and common methods include mean filling, median filling, interpolation and so on. For example, for traffic flow data, if the flow data of a certain period is missing, the average flow value of the neighboring periods before and after the period can be used to fill in; for vehicle speed data, the speed of the missing moment can be estimated by linear interpolation based on the speed value of the neighboring moments.

Standardization of data

Since the format and dimension of data collected by different acquisition devices may be different, in order to facilitate subsequent data analysis and processing, data needs to be standardized. The purpose of data standardization is to convert data from different sources into a unified format and dimension, so that the data can be comparable and operable. For example, the speed data collected by different devices are uniformly converted into kilometers per hour (km/h), and the traffic flow data is uniformly converted into the number of vehicles per hour (vehicle/h). The commonly used standardization methods include minimum maximum standardization, Z-score standardization, etc. Min Max standardization is to scale the data to the interval [0,1], the formula is: X standardization=X max − X min X − X min, where X is the original data, X min and X max are the minimum and maximum values in the data respectively; Z - Fraction standardization is to convert data into standard normal distribution, and the formula is: X standardization=σ X − μ, where X is the original data, μ is the mean value of the data, and σ is the standard deviation of the data.

Data conversion

Feature extraction: extracting useful features from raw data to better characterize traffic flow. For example, from the image data collected by the traffic camera to extract the vehicle's contour features, color features, etc., for vehicle classification and identification; from the geomagnetic sensor and microwave radar data collected to extract the vehicle's speed characteristics, flow characteristics, etc., for the analysis and prediction of traffic flow. Feature extraction can be realized through image processing algorithms, signal processing algorithms and other technologies, the extracted features can more intuitively reflect the change rule of traffic flow, to provide strong support for the subsequent data analysis and model building.

Data format conversion: converting collected data into a format suitable for analysis and processing. For example, the image data collected by the traffic camera is converted to gray-scale image or binary image for image processing and analysis; the data collected by the geomagnetic sensor and microwave radar is converted to time series data for time series analysis of traffic flow. The purpose of data format conversion is to make the data easier to process and analyze, and to improve the efficiency and accuracy of data processing.

3.2 Traffic flow monitoring and analysis

3.2.1 Real-time monitoring of traffic flow using big data technology

Data Acquisition and Transmission: Real-time traffic flow data are collected by various sensors installed on the road (e. g., traffic cameras, geomagnetic sensors, microwave radar, etc.), and the collected data are transmitted to the data center through wired or wireless networks. For example, the image data collected by traffic cameras can be transmitted to the data center through fiber optic network in real time, and the data collected by geomagnetic sensors and microwave radar can be transmitted to the data center through wireless communication modules (e. g., 4G, 5G). In the process of data transmission, it is necessary to ensure the integrity and real-time data to avoid data loss and delay.

Data processing and analysis: In the data center, the collected traffic flow data are processed and analyzed in real time. First, the data are cleaned and pre-processed to remove noise and fill in missing values to ensure the quality of the data. Then, data mining algorithms and machine learning algorithms are used to analyze the data and extract the characteristics of traffic flow, such as traffic volume, speed, occupancy rate, etc. For example, clustering can be used to analyze the traffic flow data.

For example, we can use cluster analysis algorithm to cluster the traffic flow data, and divide the traffic flow into peak hour and non-peak hour, in order to better understand the change rule of traffic flow; we can use time series analysis algorithm to model and predict the traffic flow data, predict the trend of the traffic flow in advance, and provide a basis for the traffic signal control and traffic flow optimization.

Real-time monitoring system: build a real-time traffic flow monitoring system to display real-time data and analysis results of traffic flow through visualization interface. For example, in the monitoring center of the traffic management department, the real-time image of traffic flow, traffic flow curve, speed distribution map and other information is displayed on the big screen, so that the traffic management personnel can intuitively understand the current traffic situation.

At the same time, the real-time monitoring system can also set the alarm threshold, when the traffic flow exceeds the preset threshold, the system automatically sends out an alarm to remind the traffic management personnel to take timely measures to deal with traffic congestion.

3.2.2 Analyzing the spatial and temporal distribution patterns of traffic flow data

Time distribution pattern: Traffic flow shows obvious fluctuation pattern at different times of the day. Usually in the morning peak (7: 00 - 9: 00) and evening peak (17: 00 - 19: 00), the traffic flow is larger, the speed is slower, and the congestion is more serious; while in the peak hours (9:00 - 17:00 and 19:00 - 24:00), the traffic flow is relatively small, the speed is faster, and the road capacity is relatively good. Through the time series analysis of traffic flow data, the time distribution model of traffic flow can be established, such as ARIMA model (autoregressive integral sliding average model), used to predict the traffic flow trend at a certain time in the future. For example, according to the historical traffic flow data, the establishment of ARIMA model, predicting tomorrow's morning rush hour traffic flow, traffic signal control and traffic flow optimization to provide a reference basis.

4 Experimental evaluation

4.1 Experimental design

4.1.1 Background and purpose of the experiment

With the acceleration of urbanization, urban traffic flow management is facing increasingly severe challenges. The traditional traffic management methods are unable to cope with the complex and changing traffic conditions. The emergence of big data technology provides new ideas and methods for intelligent traffic flow management. This experiment aims to verify the effectiveness of big data technology in improving the efficiency of traffic flow management through the practical application of big data technology, so as to provide a scientific basis for the construction and optimization of intelligent transportation system.

4.1.2 Experimental subjects and data collection

A typical traffic area in a large city is selected as the experimental object, which contains several traffic intersections and road sections with high traffic flow and complex traffic conditions. The experimental period is from January 2024 to June 2024, which is divided into two phases: pre-experiment (January to March) and post-experiment (April to June).

Data Acquisition Method.

Traffic flow data: Real-time collection of traffic flow, speed, occupancy and other data through traffic cameras, geomagnetic sensors and microwave radar installed at intersections and road sections.

Traffic Signal Control Data: Records traffic signal timing parameters, phase difference and other information.

Vehicle delay and queue length data:Calculate vehicle delays and queue lengths from video analytics and sensor data.

4.1.3 Experimental program design

Experimental and control group setup.

Experimental group:Applying big data technology for traffic flow management and signal control optimization.

Control group: Keeping the traditional traffic management method unchanged as a benchmark for comparison.

Assessment indicators.

Intersection capacity: Measured by counting the number of vehicles passing through an intersection per unit of time.

Vehicle delay time:Calculated by analyzing the time difference between vehicles passing through the intersection.

Queue length:Calculate the length of the vehicle queue by using sensors and video analytics.

Traffic congestion index: A traffic congestion index is calculated by taking into account factors such as speed, volume and delay time.

4.2 Experimental results and analysis

4.2.1 Presentation of experimental data

The changes in the experimental data are visualized through graphs and charts. The following are the comparison results of some key indicators.

Intersection capacity.

Before the experiment:The average capacity was 1,200 vehicles per hour.

Post-experiment: Average capacity increased to 1,500 vehicles per hour, a 25% improvement.

Vehicle delays.

Before the experiment:The average delay was 30 seconds per vehicle.

Post-experiment: The average delay was reduced to 15 seconds per vehicle, a 50 per cent reduction.

Queue length.

Before the experiment:The average queue length was 50 meters.

Post-experiment: The average queue length was reduced to 30 meters, a 40% reduction.

Traffic Congestion Index.

Before the experiment:The average congestion index was 1.5 (mild congestion).

After the experiment:The average congestion index was reduced to 1.0 (smooth).

4.2.2 Statistical analysis

The experimental results were analyzed in depth using statistical analysis methods. The t-test and analysis of variance (ANOVA) were used to verify the significant differences between the experimental group and the control group. The results show that the experimental group is significantly better than the control group in terms of intersection capacity, vehicle delay time and queue length ($p < 0.05$), which indicates that the big data technology has a significant effect in improving the efficiency of traffic flow management.

5 Concluding remarks

This paper systematically explores the application of big data technology in intelligent traffic flow management through experimental evaluation and case study. The results show that big data technology can significantly improve the efficiency of traffic flow management, reduce vehicle delay time, improve intersection capacity and reduce traffic congestion. Its successful application in different cities and transportation scenarios further proves its wide application prospect and promotion value. Through continuous research and practice, big data technology will play a greater role in intelligent traffic flow management and provide strong support for the efficient operation and sustainable development of urban transportation.

References

[1] A gradient descent algorithm with modified learning rate[J]. JIANG Wenhan;JIANG Zhiman;SUN Xuelian. Journal of Changchun University of Science and Technology (Natural Science Edition),2023(06).

[2] A review of regularization methods for convolutional neural networks[J]. Chen Kun;Wang Anzhi. Computer Application Research,2024(04).

[3] A pavement marking line extraction method based on point density and Kalman filtering[J]. LI Xiaoyu; ZHOU Mei; WANG Jinhu; YAO Qiangqiang. Journal of Chinese Academy of Sciences University,2023(06).

[4] A non-conflict machine learning prediction method for unsignalized circular intersection machines[J]. REN Lili; WU Jiangling; GUO Xuliang; ZHANG Xinyue; JIANG Tao. Science, Technology and Engineering,2023(31).

[5] Research progress of principal component analysis in the field of pattern recognition[J]. WANG Wenjie; SHI Jingshen; JIANG Nianzu; LIN Shuainan;ZHAO Rui. Journal of Baicheng Normal College,2023(05).

[6] Temperature field reconstruction based on PCA dimensionality reduction and iterative regularization[J]. BAI Yun; YAN Hua; WEI Yuan-kun. Automation and Instrumentation,2023(09).

[7] Technical analysis of image recognition based on convolutional neural network[J]. Zhu Junyu. Yangtze River Information and Communication, 2023(08).

[8] Simulation of automatic identification of traffic congestion roadway information under deep learning[J]. Cui Suping; Bu Chongyang. Computer Simulation,2023(07).

[9] A study on vehicle trajectory prediction with hybrid instructional long and short-term memory networks[J]. FANG Huazhen; LIU Li; XIAO Xiaofeng; GU Qing; MENG Yu. Transportation Systems Engineering and Information,2023(04).

[10] An airspace traffic posture identification method based on improved support vector machine[J]. ZHU Chengyuan; ZHANG Che; GUAN Jianhua. Transportation Information and Safety,2023(02).