Research on Intelligent Prediction of Urban Short-term Traffic Flow Based on Big Data

Haixiang Lang1,*
1Shanxi Police College, China, 030031

*Corresponding author e-mail: 516174398@qq.com

Abstract. Traffic congestion has been the scale of major cities. Urban highway is an important carrier, which is to ensure the normal operation of the whole city. Therefore, we must establish intelligent transportation system, which is the most effective measure for intelligent prediction (hereinafter referred to as IP) of urban short-term traffic flow (hereinafter referred to as STTF). Among them, STTF prediction and big data are the key technologies, which require us to improve the accuracy and real-time performance of the prediction. This prediction model and intelligent system can be realized according to the idea of IP. The existing STTF prediction research is often based on the shallow model method, which can’t fully reflect the characteristics of STTF. Based on the IP method, this paper proposes a prediction model based on the distributed processing framework of MapReduce and BP neural network in Hadoop environment. Finally, this paper analyzes the basic process of the two runtime.

Keywords: STTF, Big Data, IP

1. Introduction

With the improvement of urbanization level, the urban motor vehicles are increasing rapidly. Therefore, the serious imbalance of traffic supply and demand will lead to increasingly serious urban traffic congestion problems, such as driving difficulties, parking difficulties, increasing traffic accidents and so on [1]. The main way is to develop intelligent transportation, which is an IP system. Through real-time, we can solve the problem of urban congestion the most effective way, which is also the future direction of transportation [2]. According to the lifeblood of intelligent transportation, the big data of real-time road observation is the basis of traffic flow prediction (hereinafter referred to as TFP), and big data provides the possibility of TFP only. At present, the performance price ratio of traffic information collection and transmission equipment is becoming more and more reasonable, and the accuracy, breadth and content of traffic data have been continuously improved [3].

Under traffic big data, by mining the traffic characteristics analysis and TFP methods of big data technology, we can get real-time traffic information, which will master the traffic flow characteristics...
and change rules. Through the establishment of a more accurate prediction model, we can make timely and accurate prediction of future STTF changes, which will effectively improve the traffic congestion. Therefore, the era of transportation big data enriches the theoretical system of urban road traffic, which improves the overall application of intelligent transportation \cite{4}. In the context of transportation big data, we should study the change characteristics and evolution law of transportation system through the Internet of things (hereinafter referred to as IOT), cloud computing, big data and other technologies, which will better develop and study the TFP method driven by big data. Based on the traffic big data to detect the law of traffic evolution, we can predict and warn the short-term change state of traffic flow, which can improve the capacity of urban road network \cite{5}.

2. Challenges of short term traffic flow intelligent prediction method

2.1. Traffic vulnerability

Traffic vulnerability refers to the phenomenon that when the traffic network is disturbed, the service level of the whole transportation system drops obviously. With the expansion of urban roads, the topological relationship of modern urban road network is becoming more and more complex, which will lead to the aggravation of the impact between node traffic flows. With the rapid increase of traffic demand, the traffic volatility will increase, which will reduce the ability of traffic system. Traffic vulnerability will increase and make the whole transportation system and its sensitivity. Therefore, the slight disturbance will make the traffic flow change greatly, which will make the prediction more difficult \cite{6}.

2.2. Traffic modeling

The theoretical prediction of linear and nonlinear systems is based on mathematical formula and flow data, which does not consider the characteristics of flow evolution. Based on linear system theory, traffic flow intelligent prediction method (hereinafter referred to as IPM) is not suitable for STTF with short time interval and large uncertainty, which is more suitable for medium and long-term and STTF prediction. Even by combining the models, we can improve the prediction accuracy, which will increase the complexity of the model.

2.3. Prediction accuracy

The traffic congestion aggravation will lead to higher and higher requirements for traffic control. Accurate TFP results are the data support of urban traffic fine control and guidance, which makes the prediction need higher precision and more efficient principles. Through observation and collection, we can make online prediction when the traffic volume is large, which is difficult to meet the accuracy requirements. Complex algorithm also makes prediction response not timely, which will lead to poor portability. Facing the new demand of IPM, this paper proposes a new idea of STTP in big data environment.

3. The prediction method of Short term traffic flow

The measurement period can range from a few minutes to a few hours, which will use data, including current observations, historical values and so on. Random factors have great influence on STTF, which will lead to strong uncertainty. The IPM should have real-time, accuracy, robustness, adaptability and
portability. According to the characteristics of the model, IP methods can be classified into the following four categories, as shown in Figure 1.

| Based on linear system theory | Based on nonlinear system theory | Intelligent prediction based on Knowledge Discovery | Combination forecasting model |
|--------------------------------|----------------------------------|-----------------------------------------------|----------------------------------|
| • Historical average model    | • Wavelet analysis model         | • Support vector machine                      | • Data aggregation and decomposition |
| • Time series model           | • Prediction method based on Catastrophe Theory | • Neural network prediction model             | • Fusion of prediction results     |
| • Kalman filtering            | • Prediction method based on Chaos Theory | • Nonparametric regression prediction model    |                                  |
| • Linear regression model     |                                  |                                                |                                  |

**Figure 1.** Short term IPM of traffic flow

Short term flow IPM has been studied for a long time, which is difficult to obtain the early traffic data. So, the data is small in volume and low in quality, which leads to the prediction research though the small sample data. Therefore, the pure mathematical theory of complex models are introduced into the prediction research, which can better carry out the inherent characteristics of traffic flow.

4. **Research on Key Technologies of IP based on big data**

4.1. *Hadoop big data platform*

Hadoop is an open source distributed computing framework, which is derived from the development of Apache Lucene. It can provide a stable and reliable interface for clusters composed of a large number of cheap hardware. The core components of Hadoop are HDFS and MapReduce, as shown in Figure 2. At present, Hadoop has become the mainstream technology of non-relational big data, which is a more extensive cloud computing platform.

**Figure 2.** Hadoop ecosystem

4.2. *Spark real time computing platform*

Spark is a big data distributed computing framework, which is one of Apache open source projects. Spark can provide a general data processing platform, which has become the fastest open source engine
for big data processing. Through the API, we can make each processing part integrate seamlessly in memory, which will cooperate to complete the overall task of the system.

4.3. BP algorithm flow

BP algorithm flow is shown in Figure 3.

![Figure 3. BP algorithm flow](image)

4.4. MapReduce implementation of BP algorithm

Combined with the algorithm flow chart, this paper analyzes the implementation process of map function and reduce function based on BP algorithm, as shown in Figure 4.

![Figure 4. MapReduce implementation of BP algorithm](image)

5. Conclusion
With the development of science and technology, IOT makes people more and more urgent for the surrounding real-time traffic information, which puts forward a new research direction in the field of intelligent transportation, such as truthful timeliness, interactivity, mobility, service, etc. The Internet technology based on cloud computing platform can meet people's needs, which makes the research of cloud platform application service more important. Short term TFP has always been a hot spot in the field of intelligent transportation, which is the basis of its application service. Through BP neural network algorithm, we can forecast STTF. Then the algorithm is implemented on MapReduce platform. Through the distributed characteristics of cloud platform, we can ensure the prediction accuracy of the algorithm, which can improve the computational efficiency of the algorithm.

**Acknowledgments**

This work was financially supported by Research on short-term prediction of urban traffic flow based on big data, Project No. 2019yyb001.

**References**

[1] Hai Ting. Research progress of IPMs on the STTF [J]. Transportation engineering and information, 2019 (4): 84-91.

[2] Wu Bo, Xu Zhiguang, Wang Feng, et al. Architecture and key technologies of intelligent transportation system based on sensor networks [J]. Highway transportation technology, 2014 (3): 126-130.

[3] Guo Haifeng, Fang Liangjun, Yu Li. Short term TFP method based on Fuzzy Kalman filter [J]. Zhejiang University of technology, 2013, 41 (2): 218-221.

[4] Lu Haiting. Research progress of STTF IPMs [J]. The transportation engineering and information, 2019, 7 (4): 84-91.

[5] Fu Gui. Short term TFP model based on support vector machine regression [J]. South China University of Technology, 2013, 41 (9): 71-76.

[6] Man Ruijun, Liang Xuechun. TFP based on multiscale wavelet support vector machine [J]. Computer simulation, 2013, 30 (11): 156-159.