

Exploration and Research on Intelligent Traffic Control System based on Cloud Computing

Hongxia Liu

Department of Electronic Information Engineering, Handan Polytechnic College, Handan, 056001, China

hongxiali2014@yeah.net

Keywords: Intelligent transportation; cloud computing; MapReduce; traffic flow forecasting

Abstract. As an important part of the intelligent transportation system, the short-term traffic flow forecasting is always one of hot research. The short-time traffic flow prediction based on MapReduce technology is put forward in the paper from the view of improving practical prediction algorithm of cloud computing platform and in combination with the technical advantage in mass data storage and massively parallel real-time processing. Improve the computational efficiency of the short-time traffic flow forecasting under the premise of ensuring the accuracy of prediction so as to enhancing the practical of the prediction algorithm. Simulation results show that the problems of long training time of the early network weights in BP neural network algorithm and of low efficiency for matching the searching history database in K nearest neighbor nonparametric regression algorithm pattern. And, it can effectively improve the prediction of the practicability of the algorithm.

Interface

Cloud computing is a new computing patterns developed in recent years. It integrates all software, hardware and information resources based on the existing internet to form new system architecture to provide services and to provide intelligent information service as low cost and high efficiency needed by people at all levels[1-2]. The internet resources service provided by cloud computing is called "cloud". Cloud computing has become a new industry with a huge potential and broad market prospects recognized in the world. At present, investment on cloud computing has been carried out around China[3-4].

Some domestic companies as Baidu, Alibaba, Grand and Telecommunication Operators are also building their own cloud computing platform as the company's data processing and storage platform and commercial application[5]. These commercial companies have brought new hope for promotion of cloud computing technology development, especially the complex and changeable traffic system area.

Traffic data preprocessing

Features of traffic data: Various data types: The acquisition terminals for traffic data comprise a coil detector, infrared detector, ultrasonic detector, microwave detector, video detector, taxi GPS and floating car. Each kind of acquisition terminal will feedback different data format to the data center.

A large amount of data: In general, the acquisition terminals for traffic data are distributed in many road of city network and transmit the information to the data center after sampling on the traffic information at the locations. It achieves a comprehensive monitoring on city road traffic conditions, traffic flow information and illegal activities. However, the data mountain is higher.

The amount of data distribution is uneven: People are involved in road traffic. And people have their own preferences on travel time, travel route and travel mode resulting in the features as randomness, uncertainty, changes greatly with time and regional strong relevance of the traffic flow. It could be seen that the distribution of data amount acquired by the terminals are uneven.

Traffic data preprocessing method: High quality traffic data is a prerequisite factor for traffic flow predictive in the intelligent transportation system. Data quality will determine the accuracy of traffic flow prediction. Therefore, before performing the traffic flow prediction, it is very important to process the traffic data. Common pretreatment methods of traffic data are including data integration, data cleaning, data conversion and data protocol.

Experimental data preparation: The data in the paper was sourced from the acquired traffic data in the California Expressway USA performance evaluation system (PeMS). The sampling time was from Oct. 1st, 2010 to Oct. 31st, 2010. Sampling time interval was 5min. Data format is as shown in table 1 and all data description in table 1 is as shown in table 2.

Table 1. measured traffic data format

Timestamp	Station	District	Route	Direction	Station Type	Samples	observed	Total Flow	Avg Occupancy	Avg Speed
10/01/2010 00:00:00	311831	3	5	S	OR	9	0		0.0336	51.7
10/01/2010 00:00:00	311832	3	5	S	FR	9	100	8	0.0241	
10/01/2010 00:00:00	311844	3	5	N	OR	20	100	2		
10/01/2010 00:00:00	311845	3	5	N	FR	10	100	4	0.0779	68.7
10/01/2010 00:00:00	311847	3	5	N	OR	30	100	9	0.0133	69.1
10/01/2010 00:00:00	311864	3	5	N	FR	10	100	2		
10/01/2010 00:00:00	311903	3	50	E	ML	24	100	105		

Table 2. Data description

Serial No	Data Item	Describe
1	Timestamp	Time stamp
2	Station	Detector identification
3	District	The detector area
4	Route	Itinerary
5	Direction	The direction of travel
6	Station Type	FR= Downhill journey; OR= Upward slope; ML= Arterial road
7	Samples	The number of samples from all lanes
8	Observed	Road cycling percentage of the total number of observation point Lane

9	Total Flow	5 minutes after the traffic lane of the sum of all observation points
10	Avg Occupancy	The average share of all lanes of a observation points in 5 minutes
11	Avg Speed	Average weighted all vehicles speed observation point 5 minutes

The traffic flow data is fitted. Although it can repair the loss and abnormal data to a certain extent the noise pollution data still exists. De-noising and smoothing the traffic data by filtering method. There are lots of functions in matlab toolbox for de-noising and smoothing. The function smooth is adopted for the smoothing time series in the paper and the smooth function provides more smoothing processing function items. For example, acquiescent moving average method, lowess, rlowess, loess, sgolay. Take the corresponding data of 311903 road detector of the main road as example, the contrasting results before and after treatment are shown in Figure 1.

The function for normalized treatment of sample data shall be:

$$\bar{x}_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$$

The formula for the data is updated after executing the algorithm:

$$x_i = \bar{x}_i (x_{\max} - x_{\min}) + x_{\min}$$

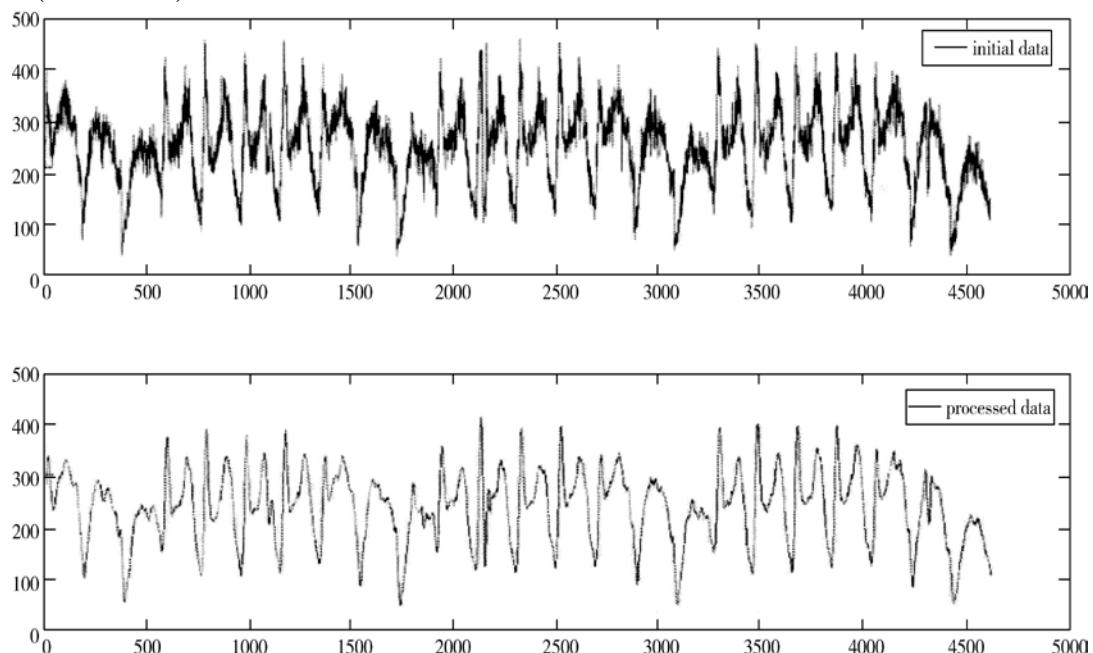


Fig. 1 Corresponding sample of 311903 detectors

Neural network algorithm for prediction of short-term traffic flow based on MapReduce

Open source platform for cloud computing – Hadoop: Hadoop is a distributed system infrastructure and it is developed by the Apache foundation. It is the developed open source platform according to GFS distributed file system of Google (Google File System) and MapReduce distributed programming model. Hadoop consists of distributed file system HDFS (Hadoop Distributed File System) component and the MapReduce programming model.

HDFS distributed file system can have a large data set (Large Data Set) applications to improve high throughput data access support, highly fault-tolerant mechanism and access the data in the file system in the form of streams and it solves the data access speed and security problem.

GA-BP algorithm design and implementation based on MapReduce: The paper will give introduction in detail in combination with the realization process of Map and Reduce function for GA-BP algorithm. The specific steps of the Map phase task allocation are as follows:

Receive sample data according to the input component and the expected output value from decomposition of the network structure. Read the weights and the threshold from the global neural

network on the distributed file system HDFS of Hadoop.

Update the local neural network weights and threshold to get result that the work signal propagates to positive output. Judge the actual output and expected output value error. If the error does not meet the requirements on training times, step shall be performed. Otherwise, Map algorithm is ended and output the result.

Prediction of short-term traffic flow by K nearest neighbor algorithm based on MapReduce

Prediction of short-term traffic flow by K nearest neighbor algorithm: We could predict the future traffic flow state through the historical changes in traffic flows model due to that the traffic flow changes is not completely random and there are repeated patterns. Predict the short-term traffic flow by KNN algorithm, as shown in fig. 5, the traffic data sampled shall be treated first to establish the historical database and find out the point that mostly matches the history data of K group by searching the history sample database and finally predict the algorithm to calculate the traffic flow of next time.

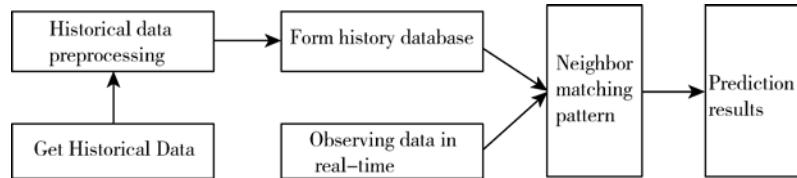


Fig. 5 Prediction process for K nearest neighbor nonparametric regression method

Experiment: In the paper, the open source Hadoop technology is used to build an experimental platform which has a total of ten machines in the machine cluster on the experimental platform. Each node is connected to the Ethernet switch of 100Mbps. Configuration of PC machine in the cluster is the same as following: dual core AMD clon 2212 2.0GHz processor, 2GB rams, 80G SCSI hard drives, Inter82551 10/100Mbps Ethernet controller; and the software environment is the Ubiintulo server operating system. Hadoop 20.2 of Cloudera is used to build a cloud computing platform and algorithm of the compiler environment is Eclipse 3.6.2, JDK1.6.0-24. Single NameNode and JobTracker are running on the primary server. It runs a DataNode and a TaskTracker from each node. DFS block size is 64M.

In order to test the prediction accuracy and acceleration ratio of short-time traffic flow prediction for GA-BP algorithm based on MapReduce, GA-BP algorithm shall be operated under the single environment of experiment A and be operated under the cluster environment of experiment B. During the experiment, select the 5-7-1 model of neural network structure, the network initial weight is random number between -0.1 and 0.1, system error: 10.5, the genetic algorithm population size: 50, the genetic crossover rate: 0.4, the variation rate: 0.07 and the evolution algebra: 30. Number of training times is set to 100 so that the execution time of Map function is not over L. Experimental prediction results are as shown in table 3.

Table 3. Contrast of prediction results of neural network algorithm

Experimental group	Training Frequency	Training times in Map	MAPE/%	Training time /h	Speed-up Ratio
A	100		9.721	104.28667	1
	200		9.637	208.73194	1
B	1	100	9.869	7.00139	14.8951
	2	100	9.821	14.19278	14.7069

It can be seen from the table, the prediction accuracy of single GA-BP algorithm is relatively close to the prediction accuracy of GA-BP algorithm based on MapReduce. However, it can be seen from the accelerate rate volume in table 2, the computational efficiency of the latter one is promoted 14 times more than the former one. The experimental curves of the prediction results of two groups

are shown in fig. 6 and 7.

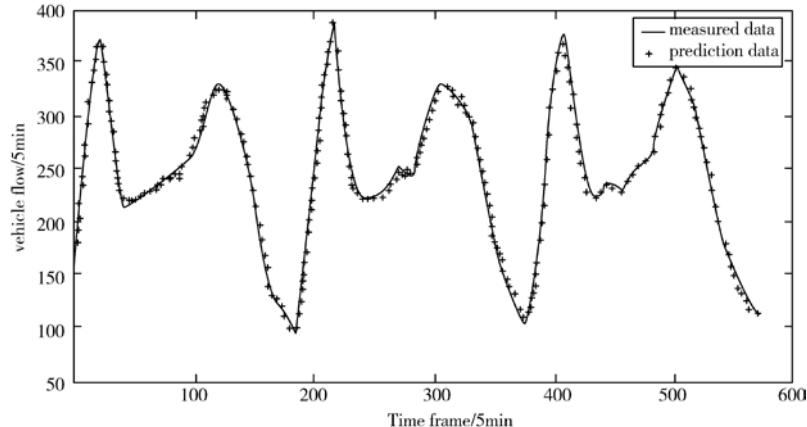


Fig. 6 Curve for the prediction results of single GA-BP algorithm

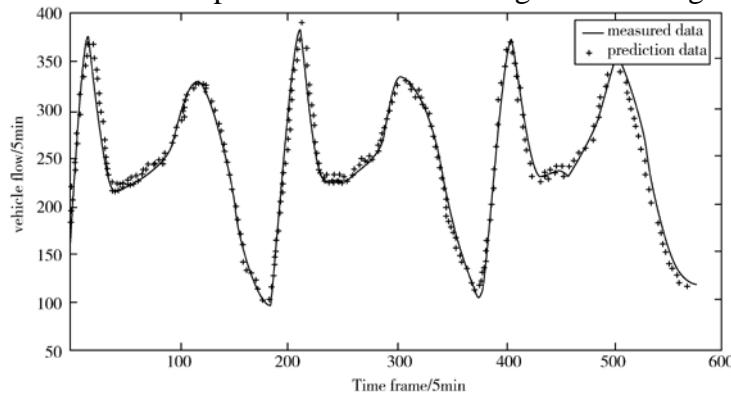


Fig. 7 Prediction results of GA-BP algorithm based on MapReduce

Conclusion

The short-time traffic flow prediction based on MapReduce technology is put forward in the paper from the view of improving practical prediction algorithm of cloud computing platform and in combination with the technical advantage in mass data storage and massively parallel real-time processing. For the adjustment and improvement of short term traffic flow of the two common short-term traffic flow prediction, improve the computational efficiency of the short-time traffic flow forecasting under the premise of ensuring the accuracy of prediction so as to enhancing the practical of the prediction algorithm.

Acknowledgments

The work was supported by the Handan Science and Technology R & D Project, its project name Research on Related Problems of Highway Information Management.

Reference

- [1]Zhang Xiaoli, Lu Hua Zeng. Non parameter turn method used in short-term traffic flow prediction [j]. Journal of Tsinghua University (NATURAL SCIENCE EDITION), 2009, 49 (9):1471-1475.
- [2]Zhang Xiaoli, He Guoguang, Lu Huapu. Nonparametric regression of short-term traffic flow prediction method based on K-neighborhood [J]. Journal of system engineering, 2009,24 (2): 178-183.
- [3] LI S S. Implementing Short-term Traffic Flow Forecasting Based on Multipoint WPRAwit MapReduce[C]. Suzhou: 2012 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), 2012:287-291.

[4] CHANG G, GE T. Comparison of Missing Data Imputation Methods for Traffic flow[C]. Changchun: 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE),2011:639-642.

[5] ZHU C M,XU X L,YAN C P. The Research of Method of Short-term Traffic Flow Forecast Based on GA-B Neural Network and chaos Theory [C]. Hangzhou:2010 2nd International Conference on Information Science and Engineering (iCISE), 2010:1617-1620.