Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jes

JES JOURNAL OF ENVIRONMENTAL SCIENCES

www.jesc.ac.cn

Predicting vehicle fuel consumption patterns using floating vehicle data

Yiman Du[1,*], Jianping Wu[1], Senyan Yang[1], Liutong Zhou[2]

1. *Department of Civil Engineering, Tsinghua University, Beijing 100084, China*
2. *Department of Civil Engineering, Columbia University, New York 10027, USA*

ARTICLE INFO

Article history:
Received 11 December 2016
Accepted 9 March 2017
Available online 18 March 2017

Keywords:
Vehicle fuel consumption
Prediction
Floating vehicle data

ABSTRACT

The status of energy consumption and air pollution in China is serious. It is important to analyze and predict the different fuel consumption of various types of vehicles under different influence factors. In order to fully describe the relationship between fuel consumption and the impact factors, massive amounts of floating vehicle data were used. The fuel consumption pattern and congestion pattern based on large samples of historical floating vehicle data were explored, drivers' information and vehicles' parameters from different group classification were probed, and the average velocity and average fuel consumption in the temporal dimension and spatial dimension were analyzed respectively. The fuel consumption forecasting model was established by using a Back Propagation Neural Network. Part of the sample set was used to train the forecasting model and the remaining part of the sample set was used as input to the forecasting model.

Introduction

Most of the metropolitan areas of the world are suffering from traffic congestion, excessive energy consumption and traffic emissions, which have become obstacles to the development of cities. The air pollution and energy wastefulness have caused residential environment deterioration and immeasurable economic loss (Zhang, 2014). The China National Science and Technology Development Plan clearly pointed out that the status of energy consumption and air pollution in China is serious, and the development of an integrated transport system was proposed to solve the problem of energy consumption and pollution, traffic safety and traffic congestion (National Outline for Medium and Long Term S&T Development 2006–2020, 2006).

The increase in motor vehicle travel and vehicle ownership is bound to further accelerate energy consumption, so that the problem of energy is becoming more prominent. It has been estimated that 90% of urban air pollution in fast-growing cities in developing countries can be attributed to vehicle emissions (Zhao and Yu, 2016; United Nations Environment Program, 2010). In addition, China's traffic pollution is much higher than in other developed countries such as Japan, due to high energy consumption, high emissions and high pollution. The daily traffic jam time in Beijing increased from 3.5 hr in 2011 to 5 hr in 2008. The daily commute time of residents in China's 15 major cities is 480 million hours higher than that in European countries. One billion yuan a day is lost due to the traffic issues in these cities (Niu, 2012). Even in developed countries, millions in economic losses are caused by traffic congestion. For example, the total economic losses due to traffic congestion in 439 US cities in 2010 was 115 billion US dollars, time waste of 4.8 billion hr. and fuel waste of 1.9 billion gallons (Schrank, 2011). In Tokyo, Japan, the annual

* *Corresponding author*. E-mail: ymducp@gmail.com (Yiman Du).

http://dx.doi.org/10.1016/j.jes.2017.03.008
1001-0742/

transportation costs have increased by 84.2 billion yen, and economic losses of up to 12 trillion yen are caused by traffic congestion (Cao, 2009).

It is apparent that effective techniques for estimating fuel consumption costs are essential in order to reduce unnecessary fuel wastage (Siami-Irdemoosa and Dindarloo, 2015). Fuel consumption is an important economic indicator. Under the premise of limited environmental capacity, the need for gradual control of fuel consumption is inevitable to protect the earth resources and improve the atmospheric environmental quality. Therefore, it is very important to analyze and predict the different fuel consumption of various types of vehicles under the influence of different factors.

Over the past decades, there have been several major categories of methods used in fuel consumption studies (Xu, 2012). One is to focus on the characteristics of the engine. Fuel consumption can be calculated based on engine data combined with the corresponding engine cycle conditions and vehicle technical parameters. The second is to use the carbon balance method to calculate the fuel consumption. The third one is building a fuel consumption forecasting model using independent variables affecting the fuel consumption, based on real vehicle tests. Accurate and reliable data for motor vehicle fuel consumption can be obtained experimentally. However, it is limited by the experimental equipment and conditions. Therefore, the forecasting model of fuel consumption is an important means for the evaluation of traffic strategies (Yan, 2012). In an urban traffic environment, there are many factors that affect the fuel consumption forecasting model, such as vehicle performance, roads, traffic conditions, vehicle age distribution, driving range, driving speed, environmental conditions, road conditions, traffic volume, speed, *etc.* So far, most of the research on fuel consumption has focused on the characteristics of automobiles, and the corresponding fuel saving technology is widely used in automobile design. However, other factors such as road conditions, traffic conditions and drivers' characteristics are seldom taken into consideration. Therefore, the study of fuel consumption under consideration of different factors is of practical significance (Gao, 2007). De Vlieger (1997) found that fuel consumption will increase by 12%–40% with aggressive driving. Rakha and Ding quantified the impact of the number of stops and driving speed on fuel consumption using the VT-Micro model to prove the relationship between vehicle fuel consumption and driving behavior (Rakha and Ding, 2003).

Influenced by various factors, fuel consumption changes over a large range with no regularity. It is difficult to predict fuel consumption by using a mathematical formula. Artificial neural networks (ANNs) employ a massive interconnection of simple processing elements that incrementally learn from their environment to capture essential linear and nonlinear trends in complex data, so that they provide reliable predictions for new situations containing even noisy and partial information (Siami-Irdemoosa and Dindarloo, 2015; Schalkoff, 1997; Haykin, 1994). ANN has strong adaptability, fault tolerance and self-learning ability, and can fit arbitrarily complex nonlinear models to multidimensional data to any desired accuracy (Smith, 1996). Neural networks have been used successfully in several studies for modeling energy consumption and exhaust emissions.

In order to fully describe the relationship between fuel consumption and impact factors, a massive floating vehicle database was used in this study. The fuel consumption forecasting model was established by using a BP neural network (Back Propagation Neural Network). Part of the sample set was used to train the forecasting model and the remaining part of the sample set was used as the input to the forecasting model.

1. Data description

In order to fully describe the relationship between fuel consumption and the impact factors, massive samples of floating vehicle data (FVD) were collected. More than 13,000 floating vehicles in Beijing were involved in this project. The collected data included the time, location, speed, driving direction, road conditions, fuel consumption, the drivers' personal information, the vehicles' basic parameters and some other related information. Data were recorded every 30 sec from June 1 to July 31, 2012. The data amounted to about 300 GB.

Based on the FVD data, the relationship between the fuel consumption and each factor was analyzed and demonstrated in Table 1 and Figs. 1 and 2. The fuel consumption is described in terms of liters per 100 km.

$$\text{Fuel consumption of a car} = \frac{\text{Total fuel consumption of a car}}{\text{Total travel distance of this car}}$$

The independent samples *t*-test can be used to see if two means are different from each other when the two samples that the means are based on were taken from different individuals who have not been matched. Therefore, the independent samples *t*-test method was used to compare the means of two independent samples for each factor, and identify whether the statistical difference between two sample groups is probably representative of real difference. The significant difference index (Sig.) between two groups for each factor is shown in Table 1, and the Sig. values are smaller than 0.05, which denotes that the fuel consumption is statistically significantly different between the two groups for each factor.

Average fuel consumption of different urban road hierarchies and different ages is shown in Fig. 1. The average fuel consumption in arterial roads and branches is higher than the other two types of roads.

Table 1 – The average fuel consumption in different factors.

Fuel consumption (l/100 km)	Gender		Transmission type		Fuel type [a]	
	Male	Female	Auto	Manual	93#	97#
Mean	8.551	8.439	8.582	8.426	8.495	8.823
Sig.	0.01		0.01		0.03	

[a] The so-called No. 93, 97 unleaded petrol means that they contain respectively, 93%, 97% of the anti-knock ability of "iso-octane", which contains respectively, 7%, 3% of the anti-knock ability of pure n-heptane.

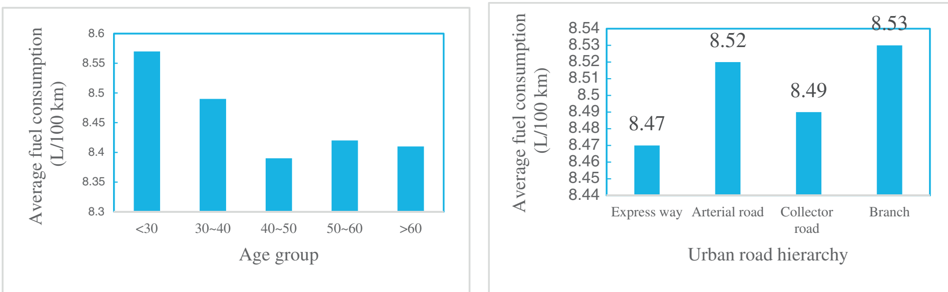


Fig. 1 – The average fuel consumption of different ages (a); and different urban road hierarchies (b).

The visualization method was adopted to match and display the fuel consumption with the road network (Fig. 2), trying to find the distribution of fuel consumption in the spatial dimension. The map is described by a two-dimensional Cartesian coordinate system, which is more complex than that in the time dimension. Baidu heat map API was used to display information on the map. The key parameter C range is 0 to 100, and can be regarded as a score for the key variables. The higher the score is, the deeper the color on the map; the lower the score, the more colorless. Fuel consumption is scored as follows: $C = \text{fuel} / \max(\text{fuel}) \times 100$.

In Fig. 2, the color gradient stands for the degree of fuel consumption. The darker the color is, the higher the fuel consumption is. Based on the map of special and temporal distribution of fuel consumption, the time and road segments connected with high fuel consumption can be found.

It was found that the proportions of male and female drivers were approximately 72% and 28%, and male drivers' driving behavior caused more fuel consumption. Automatic transmission cars accounted for a larger proportion than manual transmission and females prefer automatic transmission cars. Automatic transmission car fuel consumption is higher than that of manual cars. This result is in consistent with Wang (2002), whose experimental results reveal that the average automatic transmission vehicle fuel consumption is 0.64 L/100 km higher than that of a manual car. About 60% of the driving population is concentrated in the 30 to 45 years of age range, and drivers with age under 30 consume more fuel. About 95% of the vehicles use No. 93 fuel, but the cars using No. 97 fuel consume more fuel. The average fuel consumption during the weekday is much higher than that during the weekend based on Fig. 3. (Shang et al., 2014) analyzed the average fuel consumption around Beijing on three types of days: weekdays, weekends, and public holidays respectively.

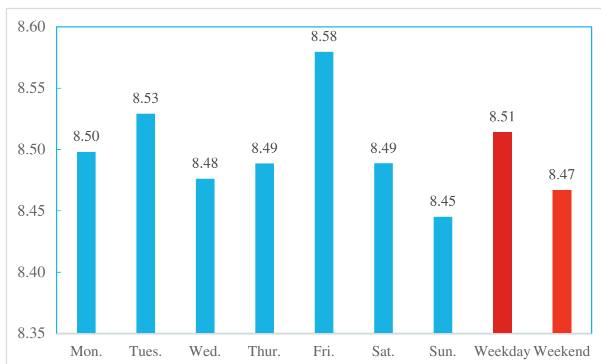
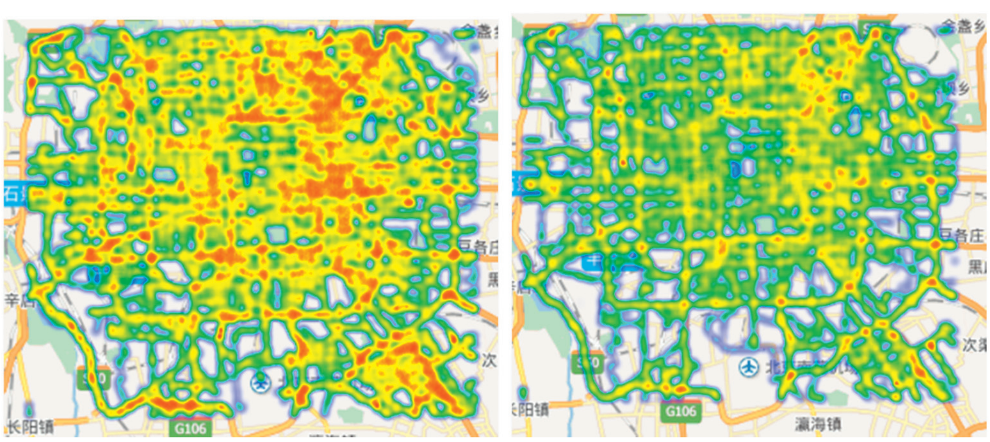


Fig. 3 – Average daily fuel consumption from Monday to Sunday.



(a) (b)

Fig. 2 – The average fuel consumption within the fifth ring road in Beijing during weekday (a) and during weekend (b).

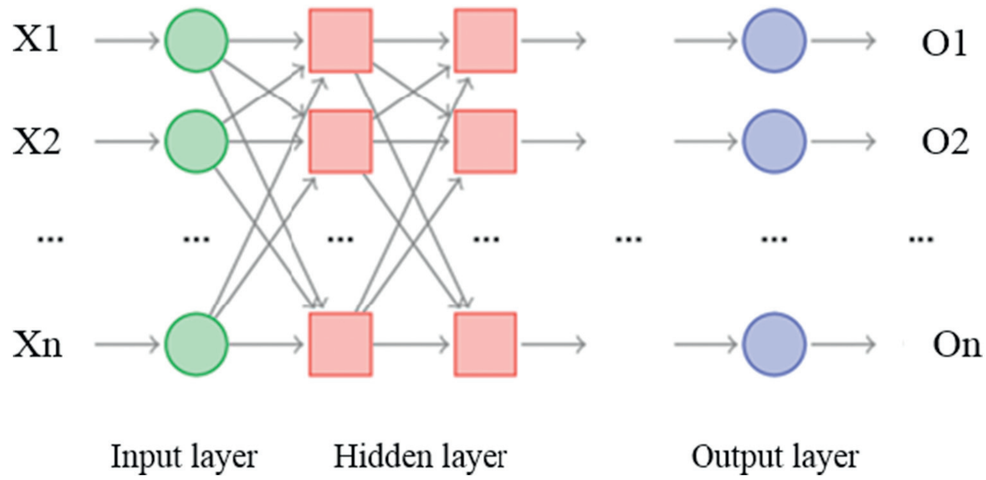


Fig. 4 – Primary structure of the proposed network.

Their analysis also showed that the fuel consumption is significantly different between weekday and weekends.

2. BP neural network based fuel consumption model

2.1. The proposed structure of the BP neural network

A BP neural network solution based on floating vehicle data is proposed in this study. The BP network consists of an input layer, an output layer, and one or more hidden layers. The primary structure of the proposed BP neural model is shown in Fig. 4. The number of neurons in the input layer is 9 and the output layer consist of one neuron. The number of hidden layers is 1. The variables of input neutrons are listed in Table 2.

2.2. The number of hidden neurons

There are some commonly used rules in calculating the minimum value of hidden neurons. Table 3 lists several of these rules. The use of 13 hidden neurons resulted in minimum MSE (mean squared error) values.

For the first two formulas, the number of hidden neurons is fixed. The trial-and-error method with increased number of hidden neurons would be better in determining the proper size of the hidden neurons (Cirak and Demirtas, 2014). Therefore, the third formula was adopted in this study starting with $\alpha = 1$. Different models with different hidden neuron numbers were conducted, and each time the number of α was increased by 2. The MSE results are listed in Table 4. It is clear that the No.4 model with 10 hidden neurons obtained the lowest MSE value (bold in Table 4). Therefore, 10 hidden neurons were adopted in this study.

2.3. The number of epochs

Different models with different epochs were trained and the results are shown in Fig. 5. It can be seen that the best solution is achieved in 2000 epochs. Acceleration is large enough to make the search path oscillate around the minimum indefinitely.

Table 3 – Some commonly used rules in calculating the minimum value of hidden neurons (Yang, 2010; Masters, 1994).

Rules	Remark
$N_h = (N_{input} \cdot N_{out})^{1/2}$	where N_{input} is the number of neurons in the input layer
$N_h = \log(2)^{Ninput}$	N_{out} is the number of neurons in output layer
$N_h = (N_{input} + N_{out})^{1/2} + \alpha$	α is the constant number between 1 and 10

Table 4 – Results of different models with different hidden neuron numbers.

No.	Model [a]	MSE
1	9-4-1	0.00032692
2	9-6-1	0.00037202
3	9-8-1	0.00019523
4	**9-10-1**	**0.00009996**
5	9-12-1	0.00025849

[a] The first number is the number of neurons in the input layer; the second number is the number of neurons in the hidden layer; the third number if the number of neurons in the output layer.

Table 2 – The variables of input neutrons.

Driver's info		Car's info				Dynamic info		
Gender	Age	Transmission type	Fuel type	Weight	Mileage	Speed	Time	Location

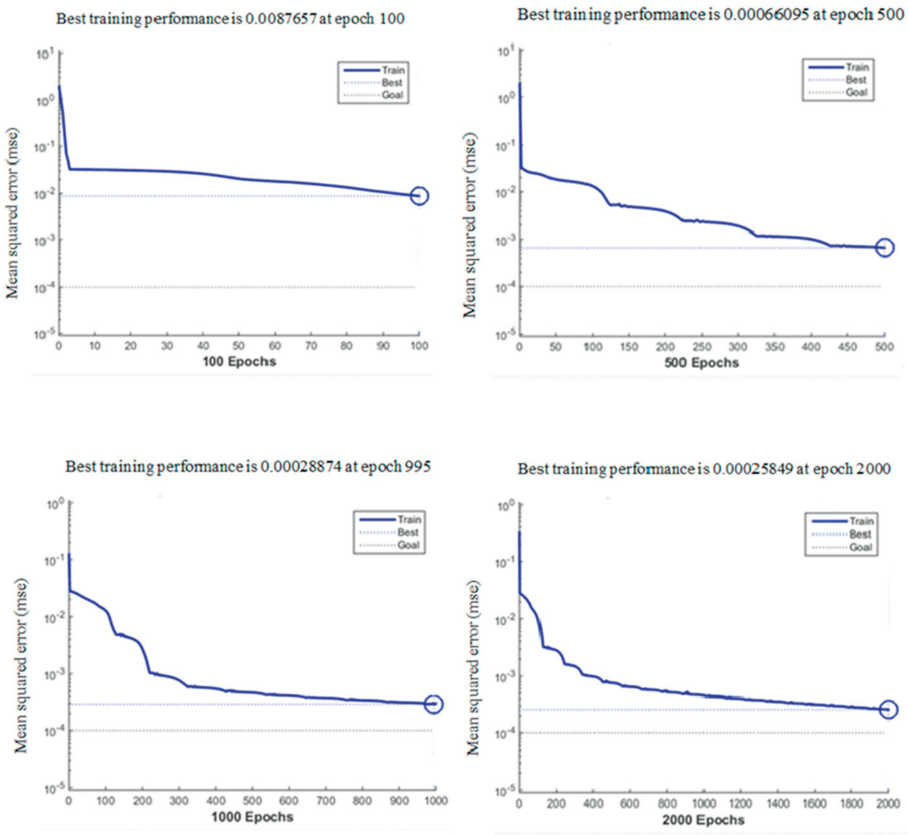


Fig. 5 – Training results for different epochs.

2.4. Results

In order to evaluate the prediction performance of BP Neural Network for the vehicle fuel consumption, the R value is selected as an indicator reflecting the correlation between the estimated value of the BP Neural Network model (outputs) and real fuel consumption value (targets). The larger the R is, the closer the correlation, and likewise, the higher the prediction accuracy is. For instance, $R = 1$ indicates that there is an exact linear relationship between outputs and targets. However, when R is close to zero, there is no linear relationship between outputs and targets. As is shown in Fig. 6, the fuel prediction accuracy of the BP Artificial Neural Network model is 0.81712, which is a rather high correlation value, implying that the BP Neural Network model is suitable for vehicle fuel prediction.

The prediction of fuel consumption for the test cycles *vs.* the actual recorded data is illustrated in Fig. 7. There exists a gap between the output and target fuel consumption for some abnormal cases with much higher actual values, which is the source of the prediction error. As a whole, the prediction of fuel consumption is quite close to the target values, which is in line with the high prediction accuracy above.

3. Conclusions

In order to fully describe the relationship between fuel consumption and the impact factors, massive amounts of floating vehicle data are used. This study especially explored fuel consumption patterns and congestion patterns based on large samples of historical FVD, probed drivers' information and vehicles' parameters from different group classification, and analyzed average velocity and average fuel consumption

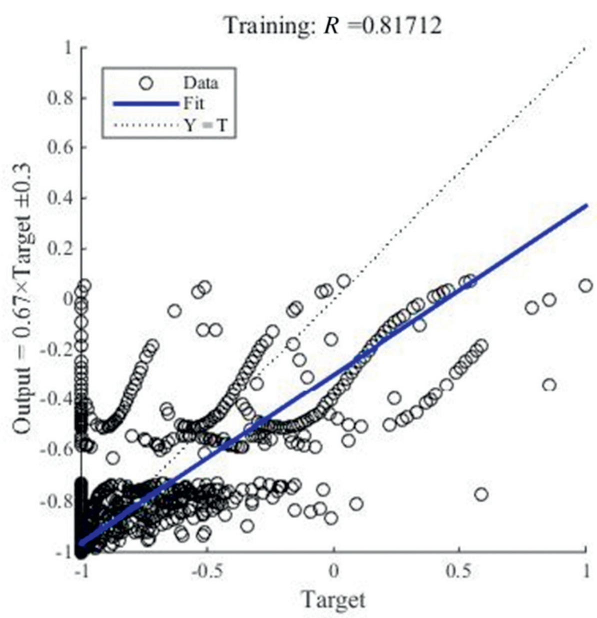


Fig. 6 – Actual *vs.* Back Propagation (BP) Artificial Neural Network model predicted fuel consumption.

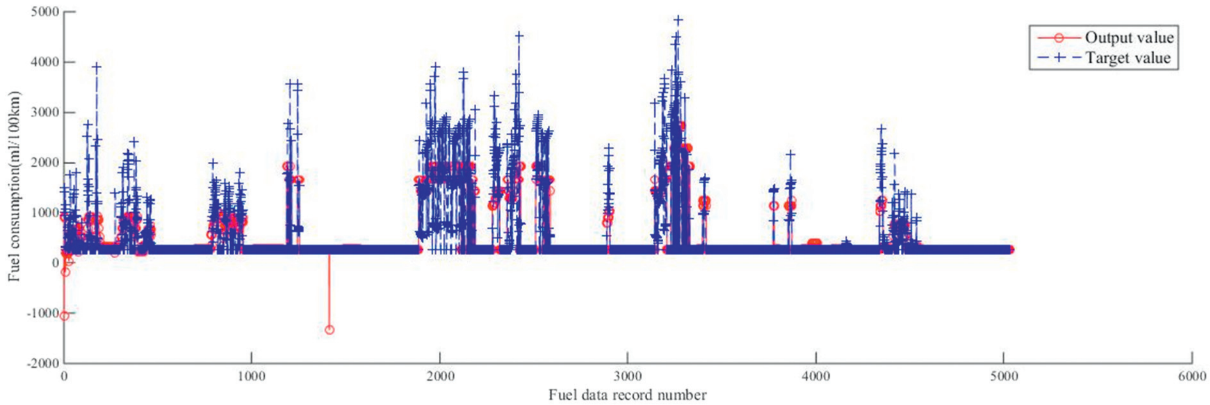


Fig. 7 – Actual *vs.* BP Artificial Neural Network model predicted fuel consumption.

in the temporal dimension and spatial dimension respectively. The fuel consumption forecasting model was established by using a Back Propagation Neural Network. Part of the sample set was used to train the forecasting model and the remaining part of the sample set was used as the input to the forecasting model.

Acknowledgments

This work was supported by the project "Research on the Traffic Environment Carrying Capacity and Feedback Gating Based Dynamic Traffic Control in Urban Network" which is funded by the China Postdoctoral Science Foundation (No. 2013M540102). The research is also supported by the Open Foundation of smart-city research center of Hangzhou Dianzi University, smart-city research center of Zhejiang Province. We also want to thank NCIC for their support in providing data and technical support.

REFERENCES

Cao, M.H., 2009. Problem in urban traffic and counter-measures analysis. J. Chongqing Jiaotong Univ. (Soc. Sci. Ed.) 9 (1), 12–15.
Cirak, Bekir, Demirtas, Selman, 2014. An application of artificial neural network for predicting engine torque in a biodiesel engine. Am. J. Energy Res. 2 (4), 74–80.
De Vlieger, I., 1997. On board emission and fuel consumption measurement campaign on petrol-driven passenger cars. Atmos. Environ. 31, 3753–3761.
Gao, L., 2007. Research on Vehicle Fuel Consumption Model Based on Urban Road Operations. Jilin University, pp. 33–57.
Haykin, S., 1994. Neural Networks: a Comprehensive Foundation. Macmillan, New York.
Masters, T., 1994. Practical Neural Network Recipes in C11. Academic Press, MA, Boston.
National Outline for Medium and Long Term S&T Development (2006–2020), 2006. The State Council of the People's Republic of China. 1 pp. 1–5.
Niu, W.Y., 2012. China New Urbanization Report. China Science Publication, Beijing.
Rakha, H., Ding, Y., 2003. Impact of stops on vehicle fuel consumption and emission. J. Transp. Eng. 129, 23–32.
Schalkoff, R., 1997. Artificial Neural Networks. McGraw-Hill, New York.
Schrank, D.T.L., 2011. 2011 Urban Mobility Report. Texas Transportation Institute.
Shang, J.B., Zheng, Y., Tong, W.Z., Chang, Eric, Yu, Y., 2014. Inferring Gas Consumption and Pollution Emissions of Vehicles throughout a City. KDD.
Siami-Irdemoosa, E., Dindarloo, S.R., 2015. Prediction of fuel consumption of mining dump trucks: a neural networks approach. Appl. Energy 2015 (04), 77–84.
Smith, M., 1996. Neural Networks for Statistical Modeling. International Thompson Computer Press, London UK; Boston, MA.
United Nations Environment Program (UNEP), 2010. Urban air pollution. http://www.unep.org/urbanenvironment/issues/urbanair.asp (Date accessed: 14 December. 2010).
Wang, F., 2002. The difference of fuel consumption between automatic and manual cars. Light Veh. 7, 47–50.
Xu, J., 2012. Research on Vehicle Fuel Consumption Model and Predication Calculation Based on Driving Style and Route Feature. pp. 5–10.
Yan, R., 2012. Research on Energy Consumption Optimization Path Analysis Algorithm Considering Terrain. Nanjing Normal University, pp. 27–28.
Yang, J.D., 2010. Research on the Test Method of Fuel Consumption Based on BP Neural Network. Jilin University.
Zhang, F.F., 2014. Research on Key Knowledge of Traffic Information System Based on Probe Vehicle. Tsinghua University.
Zhao, S.P., Yu, Y., 2016. Effect of short-term regional traffic restriction on urban submicron particulate pollution. Journal of Environmental Sciences Available online 27 August.