Neurocomputing 158 (2015) 42–47

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Development of the regional freight transportation demand prediction models based on the regression analysis methods

Yandong Yang

College of Education and P.E., Bohai University, Liaoning 121013, China

ARTICLE INFO

Article history:
Received 8 November 2014
Received in revised form
6 January 2015
Accepted 25 January 2015
Communicated by Shen Yin
Available online 17 February 2015

Keywords:
Prediction model
Regional freight transportation demand
Regression analysis methods

ABSTRACT

Different prediction models based on the regression analysis methods are studied in this work and they are successfully implemented for predicting the regional freight transportation demand (RFTD). RFTD plays an important role in reflecting the economic states, such as production improvement, economic restructuring, and economic growth style. Thus, the prediction models for RFTD have been widely used in many areas, such as academic and industrial domains. In this work, based on different prediction models, several Regional Freight Transportation Demand Prediction Models (RFTDPMs) have been constructed by using Multiple Linear Regression (MLR), Non-Linear Regression (NLR), and Simple Linear Regression (SLR). According to the fitting efficiency, the simulation results show that the RFTDPM based on NLR offers superior performances in predicting RFTD compared with the other regression models. However, if the validation rates of the RFTDPMs are taken into consideration, the SLR based model outperforms the other two prediction models.

1. Introduction

Freight transportation has increased greatly over the last few decades [1,2]. It must meet the challenges in modern industry [3]. From the perspective of freight production, freight transportation is composed of external economic demand and transportation supply system [4]. According to the haul distance, the freight transportation demand can be divided into three classes, including regional freight transportation, interlocal freight transportation and transit freight transportation. The regional freight transportation is the most basic level of the social life in this region, and the development of the regional freight transportation is strongly correlated with the regional economic development [5,6]. The regional freight transportation also focuses on modern industrial applications. Thus, regional plan, especially logistics plan, is becoming even more important [7]. In general, the transportation demand is defined as the freight volume. The freight volume is formed by the interaction of transportation demand and transportation supply. Besides, the transportation demand should match well with the existing transportation capacity [8]. The characteristics of regional freight transportation demand are their regional feature, the pertinence to freight demand connotation, the phased development of freight demand, malconformation of freight demand distribution and the complexity and diversity of freight demand species [9].

In transportation engineering, freight volume forecast is the basis for formulating relevant policies, preparing the transportation development programs and daily management [10–12]. As an important indicator to reflect the freight transportation demand, the prediction research and analysis for freight volume has both practical and theoretical significance [4]. The forecast methods of the freight transportation demand can be divided into two categories: the qualitative forecast methods and the quantitative forecast methods. The qualitative forecast methods include specialistic forecast method, jury of executive opinion method, market investigation method and Delphi method. The specialistic forecast method is a commonly used method which is time-effective, simpler, and easier [10]. Nevertheless, the qualitative forecast methods are less accurate and have one-sideness. On the other hand, the quantitative forecast methods include time series smoothness method, regression forecast method and grey forecasting method, etc. In addition, scholars introduce artificial neural network (ANN) and an intelligent fuzzy regression algorithm into the freight transportation forecast [13–16]. Furthermore, partial least squares regression (PLS) method plays an important role in prediction [17]. A multinomial probit model with spatially and temporally correlated error structure is proposed for freight demand analysis in tactical or operational planning applications [18].

In this paper, the regression forecast methods are used to forecast the freight transportation demand based on the freight volume data set, the gross domestic product (GDP) data set and the fixed investment data set. Regression analysis methods are fit for short term and mid-term forecast. They have been applied in many fields such as analysis of data groups, estimation and statistics tests, analysis

E-mail address: yandong.yang2014@gmail.com

http://dx.doi.org/10.1016/j.neucom.2015.01.069
0925-2312/

of relationships between influential factors and prediction target values, and the study of the accuracy of forecasting results [19,9,20]. In this paper, the fitted freight transportation model is constructed based on regression analysis methods and then these methods help us to predict future trend of the freight volume. We make a comparison among the models established by using multiple linear regression (MLR), nonlinear regression (NLR) and simple linear regression (SLR) and then the best model is selected out, in order to improve the model accuracy. The methods are superior to the qualitative forecast methods in the facet of freight volume prediction. Based on the proposed methods, we can forecast the future value and error together with the accuracy of forecast result by analysing the mathematical relations among variables and the relationship between the influential factors and prediction target values [7]. The models built based on regression analysis can be used in analyzing various freight transportation issues. It also offers suggestions for future research in freight transport demand. This paper can be generalized and extended to analyze the trends of freight transportation market in other parts of China and possibly elsewhere. The freight of Shanghai is critical to regional economy [21]. Based on the data they utilized, we choose GDP and fixed investment as the influential factors and then build the models.

This paper is organized as follows. Section 2 reviews the regression forecast methods, including multiple linear regression (MLR), nonlinear regression (NLR), simple linear regression (SLR). In Section 3, the data sets are introduced and analyzed. Freight transportation forecast models are established based on the regression forecast methods, the models are validated and then a comparison among these models is provided. Finally, the conclusions are presented in the last section.

2. Modeling algorithms

In this section, the regression analysis methods [22] are used to solve the regional freight demand problems. SLR, MLP and NLR are introduced briefly. In statistics, the regression analysis [23] is a statistical process for estimating the relationships among variables [24]. We can use a model to simulate data,and then analyze the simulated data. This can help you design better experiments. Regression analysis is widely used for forecast [25,26]. These regression analysis methods are fit for short term and mid-term prediction [27].

2.1. Multiple linear regression (MLR)

Multiple linear regression [28] is a generalization of linear regression by considering more than one independent variable, and a specific case of general linear models formed by restricting the number of dependent variables to one. It formalizes a simultaneous statistical relation between the single continuous outcome Y and the predictor variables X_k $(k=1,2,\ldots p)$:

$$y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_p x_{ip}+\varepsilon_i.$$

where β_0 represents the intercept (the mean of Y when all $X_k=0$), and each β_k represents a slope with respect to X_k (the magnitude of change in the mean of Y when X_k is larger by one unit and all other predictors are held constant). The β_k are thus sometimes called partial regression coefficients [29,30].

- *Step* 1: Pack all response values for all observations into a n-dimensional vector called the response vector: $Y=[Y_1,Y_2,\ldots,Y_n]^T$ and pack all predictors into a $n\times(p+1)$ matrix called the design matrix:

$$X=\begin{pmatrix}1 & X_{11} & \ldots & X_{1p}\\ 1 & X_{21} & \ldots & X_{2p}\\ \vdots & \vdots & & \vdots\\ 1 & X_{n1} & \ldots & X_{np}\end{pmatrix}$$

- *Step* 2: In order to estimate $\beta=[\beta_0,\beta_1,\ldots,\beta_p]^T$, we want to minimize the sum of errors ε:

$$\begin{aligned}Q&=\min\varepsilon^T\varepsilon=\min(Y-X\beta)^T(Y-X\beta)\\&=\min\sum(Y_i-\beta_0-\beta_1X_{i,1}-\ldots,\beta_pX_{i,p})^2,\\ i&=1\ldots n.\end{aligned}$$

According to the invertibility of X^{TX}, the β can also calculated by

$$\beta=(X^TX)^{-1}X^TY.$$

- *Step* 3: Use the ε and β to form the final model:

$$Y=X\beta+\varepsilon.$$

2.2. Nonlinear regression (NLR)

Nonlinear regression is a procedure for fitting data to any selected equation. As with linear regression, nonlinear regression procedures determine values of the parameters that minimize the sum of the squares of the distances of the data points to the curve. The general nonlinear regression model is $Y=f(X,\beta)+\varepsilon$, for $Y=[Y_1,Y_2,\ldots,Y_n]^T$, $X=[X_1,X_2,\ldots,X_n]^T$, $\beta=[\beta_1,\beta_2,\ldots,\beta_n]^T$, and $\varepsilon=[\varepsilon_1,\ldots,\varepsilon_n]$. But in this paper, owing to the parabolic regression model, we need to use the equation $Y=\beta_0+\beta_1X+\beta_2X^2+\varepsilon$.

- *Step* 1: Collect the response variables and measurable variables, presented as $Y=[Y_1,Y_2,\ldots,Y_n]^T$ and $X=[X_1,X_2,\ldots,X_n]^T$.
- *Step* 2: In order to estimate $\beta=[\beta_0,\beta_1,\beta_2]$, we want to minimize the sum of errors ε:

$$\begin{aligned}Q(\beta_0,\beta_1,\beta_2)&=\min\varepsilon^T\varepsilon=\min(Y-X\beta)^T(Y-X\beta)\\&=\min\sum(Y_i-\beta_0-\beta_1X_i-\beta_2X_i^2)^2,\\ i&=1\ldots n.\end{aligned}$$

According to the invertibility of A^{TA},

$$A=\begin{pmatrix}1 & X_1 & X_1^2\\ 1 & X_2 & X_2^2\\ \vdots & \vdots & \vdots\\ 1 & X_n & X_n^2\end{pmatrix}$$

The β can also calculated by

$$\beta=(A^TA)^{-1}A^TY.$$

- *Step* 3: Use the ε and β to form the final model:

$$Y=A^T\beta+\varepsilon.$$

2.3. Simple linear regression (SLR)

The simple linear regression (SLR) is a fundamental form of the multiple linear regression. It analyzes the linear relation between a