Traffic Flow Forecasting Algorithm based on Spatio-temporal Relationship

Xia Ying, Li Mengxin, Yang Xue, Zhou Jianbo

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

Abstract
Accurate traffic flow forecasting is the foundation of intelligent transportation system for urban traffic control. Focusing on the nonlinear nature and spatio-temporal relationship of traffic flow, the forecasting algorithm based on spatio-temporal relationship is proposed. Firstly, grey relational analysis is used for spatio-temporal relationship analyzing. Secondly, neural network and regression analysis are applied for temporal and spatial prediction respectively. Finally, two prediction results are combined by making the weighted average. Experiments show that the proposed method has reasonable forecasting effects with high precision.

Keywords: Traffic Flow Forecasting, Spatio-Temporal Relationship, Grey Relational Analysis, Neural Network, Regression Analysis

1. Introduction
With the rapid progress of urbanization, various traffic problems relevant come out, such as aggravating traffic jam and frequently happened traffic accident [1]. As the most effective method of solving the traffic problems, intelligent transportation system (ITS) is getting more and more attentions. Traffic control is one of the core topics of ITS, while the key problem is how to carry out real-time and accurate traffic flow forecasts [2].

The transport system is a complex system affected by various factors [3]. Many researches have done to establish the forecasting models of traffic flow with higher precision and stability. Currently, the widely applied forecasting models consist of time series forecasting model [4][5], neural network forecasting model [6], the support vector machine (SVM) model [7][8], and so on. Time series forecasting model is easy realization and fast, but it neglects the spatio-temporal relationship of traffic flow and thus influences the forecasting accuracy. Considering the nonlinearity and uncertainty of traffic flow, neural network is introduced in traffic field, but the convergence speed and training time are not satisfied when samples are larger. Support vector machine (SVM) model draws the advantages of accuracy, fault tolerance and anti-jamming, but it is relatively complex and difficult to process the large training sample.

This paper focuses on the short-term (usually less than 15 minutes) traffic flow forecasting [9], and propose a forecasting algorithm by considering the spatio-temporal relationship of traffic data.

2. Forecasting Algorithm Design

2.1 Basic analysis methods

(1) Grey relational analysis
Grey relational analysis is used to study the relationship existing in grey system. Grey system has no clear regularity. The status and the relationships between boundaries of the system is difficult to describe accurately [10]. The traffic flow processing system is a typical grey system [11]. So we can use grey relational analysis to quantify the relationship between traffic flows. Through comparing the grey relevance degree, we can select the routes having strong correlation with the target route as the predicting factors.

(2) BP neural network
As a mathematic modeling method, BP neural network forecasting model could identify the feature of nonlinear system. It also has the strong ability of nonlinear function approximation, adaptive learning and fault tolerance. At the same time, real-time recursion training algorithm is used to dynamically adjust the system parameters to guarantee the real-time constraints [12]. The typical characteristic of traffic flow is non-linear and real-time, so it is suitable for predicting in temporal dimension.

(3) Regression analysis
Regression analysis proceeds from causal relationship, it can figure out the correlation coefficient between dependent variables and independent variables by mathematical statistics methods. The polynomial approximation can be determined through correlation coefficient. Theoretically, polynomial approximation can be applied to any trend curve with complex nonlinear function law, thereby to forecast the trend of future development. In traffic flow forecasting, the relationship between target route and relevant routes is causal. Therefore, regression analysis can be used for the traffic flow predicting of target route.
2.2 Forecasting Algorithm

(1) The sketch

The traffic volume on one route has connection with the ones in other time periods or on other routes\(^{[13]}\). In grey relational analysis, the grey relevance degree can be calculated by comparing the changing trends of traffic flow of different routes, thus the relationship between routes can be qualified. Comparing the relevance degree, select the routes which have relatively strong correlation with the target route as the predicting factors.

For temporal dimension, the developed BP neural network is applied to forecasting traffic flow. The temporal sequence of the target route is used as the sample to train the BP neural network until it is developed enough.

For spatial dimension, the regression equation can be established through regression analysis to forecasting traffic flow. In the equation, the target route’s traffic flow is the independent variable and the relevant traffic flows are dependent variables.

Finally, the temporal and spatial forecasting results are combined by making the weighted average.

(2) Predicting factors selecting based on spatio-temporal relationship analysis

Grey relational analysis is applied to analyze the spatio-temporal relationship of traffic flows, the routes have relatively strong correlation with the target route are selected as predicting factors.

Step 1: Data normalization.

Suppose there are N routes in the study area.

Step 2: At time t, the grey relational grade \( \gamma _ {t} \) between the relevant route \( S _ {r} \) and the target Route \( S _ {t} \) as follows:

\[
\gamma _ {t} = \frac{1}{n} \sum _ {t=1} ^ {n} e _ {t} (t) \tag{3}
\]

Then the spatio-temporal relational vector \( p \) can be established as follows:

\[
p = [\gamma _ {1}, \gamma _ {2}, \cdots \gamma _ {n}]\]

Step 4: Suppose \( H \) is a threshold. If \( \gamma _ {i} \geq H \), relevant route \( S _ {i} \), will be selected as the predicting factor.

(3) BP neural network prediction for temporal dimension

Traffic flow temporal sequence of target route is used as the sample to train the BP neural network repeatedly. The neural network will be trained to approach the changing function of the route’s traffic flow. Finally, the developed BP neural network can be applied to forecasting traffic flow in time dimension. An procedure is shown below.

Step 1: Sample data division.

As what is shown in Fig. 1, sample data \( X _ {i} = [X _ {i}(1), X _ {i}(2) \cdots X _ {i}(n)] \) is divided into many sections, and every section is \( n+m \) data elements in length and contains input data and output data. There is often some overlap with the sections.

Step 2: BP neural networks training

In order to train the network, for each section, the first \( n \) data elements are used as input data, and the last \( m \) data elements are used as target output data. Based on sample data trained by the neural network, the network is developed enough to forecasting traffic flow.

(4) Regression analysis prediction for spatial dimension

Through grey relational analysis, the predicting factors are selected to form a Route set \( \{ S _ {1}, S _ {2}, S _ {3}, \cdots S _ {M} \} \) which contain \( M \) routes.

Step 1: Build sample

Regard \( X _ {t}(t) \) (the traffic flow in target route \( S _ {t} \) at time \( t \)) as dependent variable \( Y \) and \( X _ {t}(t-1) \) (\( t \in [1, M] \)) (the traffic flow in route \( S _ {i} \) at time \( t-1 \)) as independent variable. They are used to establish the regression model.
Step 2: Establish regressive prediction model
Multiple linear regression analysis model is established, as the following equation:
\[ Y = X_0(t) = \beta_0 + \beta_1 X_1(t-1) + \cdots + \beta_n X_n(t-1) + \mu \] (4)

\( \beta \) and \( \sigma \) are variable-free unknown parameter. The parameters are figured out by means of nonlinear least square method. Thus, the regression equation Eq.(4) is estimated.

Step 3: Test the regression model
Variance analysis[14] and significance test[14] are done on the regression equation to test the regression model's significance and representation.

Step 4: If the regression equation is significant, the regression equation can be applied to predict \( X_0(t) \) by using \( X'(t) \). If not, return to step 2 to optimize the equation.

(5) Spatio-temporal combination
Since the prediction for temporal dimension and spatial dimension have their own limitations, for the realization of complementary advantages, we make combination forecasting according to the idea of data fusion. Aimed at the square sum of error (SSE), weight vectors can be obtained[15].

Step 1: Establish equations as follows:
\[
\begin{align*}
e_{1t} &= X_0(t) - X'(t) \\
e_{2t} &= X_0(t) - X''(t)
\end{align*}
\] (5)

\[
\min \sum_{t=1}^{n} \sum_{j=1}^{n} k_1 e_{1t}^2 + k_2 e_{2t}^2 \quad (6)
\]

Here, \( X_0(t) \) is the real traffic flow in target route \( S_0 \) at time \( t \), \( X'(t) \) is the forecasting result of BP neural network, \( X''(t) \) is the forecasting result of the regression equation. \( e_{1t} \) is the forecast error at time \( t \) in the BP neural network prediction model, and \( k_1 \) is the weight. \( e_{2t} \) is the forecast error in the regression analysis prediction model, and \( k_2 \) is the weight.

Step 2: The optimum weight coefficients \([k_1, k_2]\) can be figured out by solving the simultaneous Eq.(6).

\[
Q_t = k_1 Q_{1t} + k_2 Q_{2t} \quad (7)
\]

At time \( t \), \( Q_{1t} \) is the prediction result of the BP neural network prediction model for temporal dimension, and \( Q_{2t} \) is the combination prediction result.

3. Experiments and Analysis
For further specific analysis of short-term traffic flow forecast, we make use of real time traffic data collected by detectors in routes to verify the model[16]. Simulation environment of the experiment is Intel 2.40GHz CPU (4 cores), 4GB RAM and Windows 7 (64 bit).

(1) Comparing with real traffic data
We choose 4 days traffic volume data from 7:30 am to 16:00 pm, the time interval is 5 minutes. We achieve the establishment of short-term traffic flow forecasting algorithm based on spatio-temporal relationship by MATLAB programming.

![Fig. 2](image1.png)

Fig. 2 Forecasting result of the forecasting algorithm based on spatio-temporal relationship.

![Fig. 3](image2.png)

Fig. 3 Relative errors of the forecasting algorithm based on spatio-temporal relationship.

Fig. 2 shows that the forecasting outputs of the proposed forecasting algorithm based on spatio-temporal relationship.
relationship are closer to the real traffic flow, and it can reflect the changing trend of real traffic flows. As see in Figure 3, during peak traffic hour, fluctuation and randomness of the traffic flow is weak, so the prediction precision is the highest. At other times, the prediction precision decreases. For example, the second relative error is exceptionally high, nearly 40%. At that time, the traffic flow was decreasing sharply to 70.

(2) Comparing with other methods

ARIMA model and neural network model are typical methods for forecasting traffic flow, they are applied to compare the predictive effect with the proposed forecasting algorithm. In the experiment, there error indicators include mean absolute error (MAE), mean squared error (MSE) and mean absolute percentage error (MAPE). They are used to evaluate the prediction results. If the rate of error indicator is smaller, the prediction is more accurate9][10].

Table 1: Comparison of forecasting methods

<table>
<thead>
<tr>
<th>method</th>
<th>MAPE</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting Algorithm based on spatio-temporal relationship</td>
<td>0.0691</td>
<td>6.5417</td>
<td>81.1250</td>
</tr>
<tr>
<td>ARIMA forecasting model</td>
<td>0.0937</td>
<td>9.6250</td>
<td>149.2917</td>
</tr>
<tr>
<td>Neural network forecasting model</td>
<td>0.0795</td>
<td>8.2083</td>
<td>122.7917</td>
</tr>
</tbody>
</table>

From Table 1, we can see that comparing with the other two models, the proposed forecasting algorithm based on spatio-temporal relationship has the minimum errors, it gains higher prediction accuracy.

4. Conclusions

Given the characteristics of traffic flow, such as fluctuation, randomness and non-linear, individual forecasting models can hardly obtain satisfied result. To forecast the short-term traffic flow better, the forecasting algorithm based on spatio-temporal relationship is established. The neural network prediction is applied for temporal dimension and the regression analysis prediction is applied for spatial dimension. The relevance degree is quantified to select the predicting factors, and spatial and temporal forecasting results are combined. Experiment results indicate that the proposed forecasting algorithm based on spatio-temporal relationship can get reasonable prediction results.

References