A travel time forecasting model based on change-point detection method

Shupeng LI, Xiaoping GUANG, Yongsheng QIAN, Junwei ZENG
School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, 730070, China
Email: 13919091569@163.com

Abstract. Travel time parameters obtained from road traffic sensors data play an important role in traffic management practice. A travel time forecasting model is proposed for urban road traffic sensors data based on the method of change-point detection in this paper. The first-order differential operation is used for preprocessing over the actual loop data; a change-point detection algorithm is designed to classify the sequence of large number of travel time data items into several patterns; then a travel time forecasting model is established based on autoregressive integrated moving average (ARIMA) model. By computer simulation, different control parameters are chosen for adaptive change point search for travel time series, which is divided into several sections of similar state. Then linear weight function is used to fit travel time sequence and to forecast travel time. The results show that the model has high accuracy in travel time forecasting.

1 Introduction
Travel time prediction is based on different travel time information, It can be used to predict the travel time of a certain period in the future, or to predict the passage time by multiple paths travel time information. Travel time parameters can reflect the road traffic flow state and be used by a traffic management department to make the induction measures, providing the important basis of traffic information service. Improving the prediction accuracy of travel time parameters is very important to optimize the travel efficiency of individual and balance the usage of road resources.

Guin A and Billings D used the ARIMA model to predict travel time. In particular, Billings D verified the effectiveness of the model for short-term travel time prediction in urban road through the residual analysis and the lack of fit test. Mentioned research provides a good inspiration for this article. This paper studies a kind of change point searching algorithm, which divides the input data into different patterns. It also uses ARIMA model to fit the data from detectors on that basis. It is hoped that this method will provide a new method and idea for the prediction of travel time parameter sequence.

2 Travel Time Calculation
It is necessary to convert the coil detector data to travel time after travel time prediction using the traffic data obtained by the coil detector. Assuming that detectors are provided upstream and downstream of each road section, the velocity $v_i(t)$ of the vehicle between the road detectors $k$ and $k+1$ can be expressed as:
$$v_i(t) = V(k,d) + \frac{s_i(t) - s_k}{s_{k+1} - s_k}(V(k+1,d) - V(k,d))$$ (1)
\( s(t) = \int v(t) dt \). Since differential equations are often difficult to obtain exact solutions, an approximation is generally sought instead. The \( s(t) \) in the equation can be approximated obtained by the formula (2):

\[
\begin{align*}
 s_1(t) &= s_1^0 + \frac{V(k, d)(s_{k+1} - s_k)}{V(k+1, d) - V(k, d)} + s_{k+1} - s_k \\
 &= \frac{V(k, d) - V(k, d)}{V(k+1, d) - V(k, d)} (t - t_1^0) - 1
\end{align*}
\]

(2)

In the formula (2), \( s_1 \), \( s_{k+1} \) represents the position of the detectors \( k \) and \( k+1 \), \( V(k, d) \) represents the speed of detector \( k \) in time period \( d \), and \( s_1(t) \) represents the specific track of the vehicle during the time period \( d \) of the road section \( k \). For road sections, \( (v, t) \) is the initial state of the track of the vehicle entering the interval \( \{k, d\} \). If \( V(k+1, d) = V(k, d) \), the formula (2) can be converted to (3):

\[
 s_1(t) = s_1^0 + V(k, d)(t - t_1^0)
\]

(3)

According to the formula (1), (2), (3), the travel time of the road section is calculated as formula (4), (5).

When the velocity is high:

\[
 t^* = t_1^0 + \frac{(s_{k+1} - s_k)}{V(k+1, d) - V(k, d)} \ln \frac{V(k, d)(s_{k+1} - s_k)}{V(k+1, d) - V(k, d)} + s_{k+1} - s_k
\]

(4)

When the velocity is low:

\[
 t^* = t_1^0 + \frac{s_{k+1} - s_1^0}{V(k, d)}
\]

(5)

3 Building Model

3.1 Change point searching algorithm

The travel time can be divided into a similar period of time. It is possible to greatly reduce the overall error by segmented fitting regression the travel time in different states and summarizing the sequences.

\[
\bar{x}_{[i]} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

(6)

\[
\text{Dev}(x_{[i]}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_{[i]})^2
\]

(7)

The convex/concave wave:

\[
 x_{[j]} = (x_{j-1} + \ldots + x_j) / (t_j - t_{j-1} + 1)
\]

(8)

\[
\text{Dev}(x_{[j]}) = \frac{1}{n} \sum_{j=1}^{n} (x_i - x_{[j]})^2, j = 2, \ldots, k
\]

(9)

In the formula, \( \{t_j\} \) is the valley point at the peak curve or peak point on the valley curve.

When the observation function is greater than the previously determined sequence change point control parameter, the current search point is reserved, otherwise it is considered that the point is not true to the next search point. Observation function:

\[
T(w) = \frac{B(w, x_{[i]})}{B(w+1, x_{[i]})}, w = 0, 1, \ldots, J - 1
\]

(10)
\[ B(w, x_{j}) = \sum_{t=0}^{\infty} \text{Dev}(x_{j}, x_{j+1}) + \text{Dev}(x_{j+1}, x_{j}, x_{j+1}), \quad w = 1, \ldots, J \]

(11)

\[ \{ \beta(w) \} \text{ is the set of subscripts for the change points on the curve, and } \beta(0) = x_i = x_i. \]

The specific algorithm is as follows:

**Step1:** The travel time series is plotted as a curve, \( u \) is the number of cycles, \( w \) is the number of change points, and \( ee \) is the auxiliary variable.

Define \( u = 1, w = 1, ee = 0, \beta_j = t_j (j = 1, \ldots, k) \)

**Step2:** Select two convex waves along the timeline, denoted as \( (\beta_s - \beta_{s+1}, \beta_{s+1}) \). In a two-convex wave length, the wave deviation of the demand sequence and the value of each point in the two waves are obtained as the values of the intermediate point, and find the maximum and minimum \( \beta \);

\[ A(\beta, \beta_{s+1}, \beta_{s+1}) = \sum_{t \in \beta} (x_t - \bar{x})^2 + \sum_{r \in \beta_{s+1}} (x_r - \bar{x})^2 \]

(12)

\[ M(\beta, \beta_{s+1}, \beta_{s+1}) = \max \sum_{t \in \beta} (x_t - \bar{x})^2 + \sum_{r \in \beta_{s+1}} (x_r - \bar{x})^2 \]

(13)

\[ m(\beta, \beta_{s+1}, \beta_{s+1}) = \min \sum_{t \in \beta} (x_t - \bar{x})^2 + \sum_{r \in \beta_{s+1}} (x_r - \bar{x})^2 \]

(14)

**Step3:** Change point estimation. Let:

\[ R_e = 2A(\beta, \beta_{s+1}, \beta_{s+1}) / (M(\beta, \beta_{s+1}, \beta_{s+1}) + m(\beta, \beta_{s+1}, \beta_{s+1})) \]

(15)

\[ c_e = M(\beta, \beta_{s+1}, \beta_{s+1}) / m(\beta, \beta_{s+1}, \beta_{s+1}) \]

(16)

1. If \( M(\beta, \beta_{s+1}, \beta_{s+1}) = m(\beta, \beta_{s+1}, \beta_{s+1}) \), when \( 1 \leq c_e < 1.3 \) when the sequence changes gently, there is no need to define change points in the interval;

2. If \( R_e \leq \frac{1}{4} + 3/2(c_e + 1) \), Whether or not it meets \( t_s = \beta_{s+1} \), there is \( \beta(0) = \beta_{s+1} \);

3. If \( 1 > R_e \geq \frac{1}{4} + 3/2(c_e + 1) \), Look for the nearest peak or valley point of the two sides of \( \beta_{s+1} \), denoted as left \( \beta_{s+1} \), right \( \beta_{s+1} \). If \( x_e \in \{ \beta_{s+1}, \beta_{s+1} \} \), then \( \beta(0) = t_s \).

4. If \( R_e > 1 \), then \( \beta(0) = t_s \).

5. Judge the effectiveness of \( \beta(0) \) which meets \( 2 \) \( 3 \) \( 4 \); if \( T(w-1) - 1 > \psi \), keep the change point, or remove it.

**Step4:** If \( \beta(0) \) exists, \( u = u + 1, \beta_s = \beta(0), \beta_{s+1} = \beta_{s+1}, \beta_{s+1} = \beta_{s+1}, w = w + 1, ee = 0 \). Otherwise, \( ee = 1, \beta_s = \beta_s, \beta_{s+1} = \beta_{s+1}, \beta_{s+1} = \beta_{s+1} \).

When \( \beta_{s+1} = \beta_k \) the search is completed. The algorithm ends, or turn to step 2.

### 3.2 ARIMA Forecasting Model

ARIMA (p, d, q) model can be expressed as:

\[ \Phi(B) \nabla^d x_t = \Theta(B) \epsilon_t \]

(17)

In order to reduce the fitting residuals of the adjacent prediction points and improve the prediction accuracy, we use the weighted least squares method to complete the estimation of the unknown parameters of the ARIMA model. Assuming \( x_i \) the collected time series data, \( \omega \) is the weight of the point,
then the weighted least squares method is expressed as:

\[ x'_i = x_i \cdot \omega_i \]  

(18)

The least squares parameter estimation of the common ARIMA model parameters is performed for the transformed \((x'_1, \cdots, x'_n)\) to obtain the optimal parameter vector \( \hat{\beta}' \) of the model after weighting transformation:

\[ \hat{\beta}' = (\phi'_1, \cdots, \phi'_p, \theta'_1, \cdots, \theta'_q)' \]  

(19)

Weighted predictive formula:

\[ F_i(\hat{\beta}) = \phi'_0 x_{i-1} + \cdots + \phi'_p x_{i-p} + \varepsilon_i' - \theta'_1 \varepsilon_{i-1} - \cdots - \theta'_q \varepsilon_{i-q} \]  

(20)

After the weight is removed, the final forecast is obtained:

\[ F_i(\hat{\beta}) = F_i(\hat{\beta}) / \omega_i \]  

(21)

The residual sum of squares of the sample observations is:

\[ Q(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (x_i - F_i(\beta))^2 \]  

(22)

When estimating the unknown parameters of the model, the objective function is set to the minimum overall residual sum of the squares \( \min Q(\hat{\beta}) \), and the least squares parameter estimation vector is obtained.

4 Example of Verification

In this paper, the actual detector data of a city road is selected to verify the prediction model. The distribution of road sections and detectors is shown in Figure 1.

![Figure 1: A section and detector distribution diagram](image)

4.1 Travel time calculation

As shown in Figure 1, the vehicle travels from east to west, passing through road section 1, section 2. Wash the traffic flow data of three sections by the algorithm 1. Then, calculate the travel time of the preprocessing results by the formula (4), (5). And the travel time calculation results of road section 1 and section 2 are obtained. As shown in Figure 2.

![Figure 2: Travel time calculation result](image)
4.2 State change point search
Select 7 different control parameters \( e_0 = 0.01, 0.02, 0.05, 0.08, 0.10, 0.12, 0.14 \). And use the algorithm 2 to self-adapting search the travel time series for change point. The search results are shown in Table 1.

| \( e_0 \) | search results                | number |
|--------|------------------------------|--------|
| 0.01   | \{1, 24, 60, 69, 78, 90, 222, 320, 325, 341, 357, 377, 379, 390, 395, 415\} | 16     |
| 0.02   | \{1, 24, 90, 222, 377, 381, 385, 390, 395, 415\} | 10     |
| 0.05   | \{1, 22, 377, 381, 390, 395, 413\} | 7      |
| 0.08   | \{1, 223, 381, 395, 413\} | 5      |
| 0.10   | \{1.412\} | 2      |
| 0.12   | \{1, 416\} | 2      |
| 0.14   | \{1\} | 1      |

It can be seen from Table 1 that you can achieve the control of the state change point search results by adjusting the \( e_0 \) size. However, in the actual forecast work, too much the state change point is not much significance, so we need to choose the appropriate parameters according to the actual needs of the search process. According to the characteristics of travel time series, it is appropriate to select the search results for 5 ~ 10 parameters \( e_0 \) of the change points, so \( e_0 = 0.08 \), \( e_0 = 0.05 \), \( e_0 = 0.02 \) are reasonable control parameters. Comparing the state segmentation results, \( e_0 = 0.02 \) and \( \bar{y}_0 = \{1, 24, 90, 222, 377, 381, 385, 390, 395, 415\} \) are the ideal control parameters and change point recognition results, as shown in Figure 4. According to the search result, the travel time series is divided into 6: 00 ~ 6: 50, 6: 50 ~ 9: 00, 9: 00 ~ 13: 28, 13: 28 ~ 17: 00, 17: 00 ~ 18: 38, 18: 38 ~ 18: 56, 18: 56 ~ 19: 48, 19: 48 ~ 23: 00 a total of ten states of this similar section.

5 Calculation results and analysis
As the segmentation parameter regression can get better results than the global optimization, this article selected 13: 28 ~ 17:00 travel time sequence. ARIMA model was used to model building and parameter regression, so as to predict the travel time at 17:02 and 17:04.

Analyze the stability of the selected time series, and make sure that it is a smooth sequence. So the different order degree of the ARIMA \((p, d, q)\) model \( d = 0 \). Perform pure randomness tests on the sequence. And the sequence is determined to be a non-white noise sequence. Calculate the
autocorrelation coefficient and the partial autocorrelation coefficient of the sequence. The autocorrelation coefficients show obvious trailing characteristics, and the partial autocorrelation coefficient shows the truncation characteristic, so that the order of the ARIMA \((p, d, q)\) model \(p = 2, q = 0\) is determined. Finally, the ARIMA model results of the travel time series of 13:28 ~ 17:00 are \((2, 0, 0)\).

Using the linear weight function \(y = x\) to fit the travel time series, we get the transformed fitting function as \(x'_t = 0.000957685 + 0.626274x_{t-1} + 0.100417x_{t-2}\). The mean absolute error MAE is 0.0110502. The fitting results are shown in Figure 4.

Figure 4: Linear weight function arima model fitting results

The predicted values of travel time at 17:02 and 17:04 calculated by fitting function are 0.207165, 0.200433. Summarize the real values of the two moments travel time and the predicted values fitted by the ARIMA model. And analyse the error analysis of the fitting accuracy. The results are shown in Table 2.

| Predictive results and error analysis | 17:02 | mean absolute percentage error (MAPE) |
|--------------------------------------|------|--------------------------------------|
| actual value                         | predict value | value | 1.31% | 1.38% |
| 0.204493                             | 0.207165     | 0.200433 | 0.197695 | 0.200433 | 1.31% | 1.38% |

It can be seen from Table 2 that the average absolute percentage error of the travel time prediction result at 17:02 is 1.31%, the travel time prediction results at 17:04 average absolute percentage error of 1.38%, the forecast results can meet the prediction accuracy error within 10% of the basic requirements. The results of the error analysis of the predicted and real values show the ARIMA model with linear weight function has a good fitting effect on the experimental data and the prediction result is more accurate.

6 Conclusion
Based on the ARIMA model, the forecasting model of travel time is established by using the change point searching algorithm to pattern clustering the travel time parameter sequence. Through simulation
verification, travel time has achieved high precision prediction results. And the results show that the method can be used to forecast the urban road travel time based on the coil detector data.

Acknowledgment

This work is supported by The National Planning Office of Philosophy and Social Science (No.15BJY037), The Natural Science Foundation of China(No.51468034), The National Planning Office of Philosophy and Social Science(No.14CJY052).

References

[1] Li C S, Chen M C. “A data mining based approach for travel time prediction in freeway with non-recurrent congestion’, Neurocomputing, volume133, pp.74-83, (2014)
[2] Park H, Haghani A. “Optimal number and location of Bluetooth sensors considering stochastic travel time prediction”, Emerging Technologies, volume55, pp. 203-216, (2015)
[3] Yildirimoglu M, Geroliminis N. “Experienced travel time prediction for congested freeways”, Transportation Research Part B, volume53, pp. 45-63, (2013)
[4] Zhang X, Rice J A. “Short-term travel time prediction” Transportation Research Part C: Emerging Technologies, volume11, pp. 187-210, (2003)
[5] Chen H, Rakha H A. “Real-time travel time prediction using particle filtering with a non-explicit state-transition model”, Transportation Research Part C: Emerging Technologies, volume43, pp.112-126, (2014)
[6] Jou Y J, Wen Y H, Lee T T, et al. “Missing data treatment on travel time estimation for ATIS Systems, IEEE International Conference on2003, volume1, pp.102-107, (2003)
[7] YANG Li, YANG Deming, “Study on passenger flow change in shenzhen urban rail transit network”, Modern Urban Transit, volume6, pp.72-75, (2012)
[8] CHENG Cheng, DU Yuchuan, LIU Xin. “A passenger volume prediction model of transportation hub considering holiday effects”, Transportation Systems Engineering and Information, volume15, pp. 202-207, (2015)
[9] FAN Na, ZHAO Xiangmo, DAI Ming, et al. “Short-term traffic flow prediction model”, Journal of Traffic and Transportation Engineering, volume12, pp.114-119, (2012)
[10] WANG Jian, DENG Wei, ZHAO Jinbao. “Short-term freeway traffic flow prediction based on multiple method-s with bayesian net-work”, Transportation Systems Engineering and Information, volume11, pp.147-153, (2011)