Abnormal traffic flow data detection based on wavelet analysis

Qian Xiao[1,a], Dong Jian Wang[1] and Dan Liu[1]

[1] Institute of Information Engineering, Shenyang University, Shenyang 110044, China

Abstract. In view of the traffic flow data of non-stationary, the abnormal data detection is difficult.proposed basing on the wavelet analysis and least squares method of abnormal traffic flow data detection in this paper.First using wavelet analysis to make the traffic flow data of high frequency and low frequency component and separation, and then, combined with least square method to find abnormal points in the reconstructed signal data.Wavelet analysis and least square method, the simulation results show that using wavelet analysis of abnormal traffic flow data detection, effectively reduce the detection results of misjudgment rate and false negative rate.

1 Introduction

At present, the supply of the existing urban transportation infrastructure cannot meet the needs of the current urban development on the road, which is he main reason that caused by traffic congestion in the great cities of the world, traffic safety, and environmental pollution. Therefore, in today's rapid economic development and science and technology innovation, will have to take a new urban traffic management technology. At present, more advanced, more widely used a traffic management technology is intelligent transportation system (ITS). It through the use of advanced electronic, information, sensor and system engineering, and other technology integration and application based on relatively perfect road facilities to satisfy the actual demand of ground transportation. Traffic monitoring is through many data to rely on, general conditions have been monitor to collect traffic data. But the weather situation, equipment failure, traffic control and other reasons will often cause the lack of data acquisition and error phenomenon. Therefore, the abnormal traffic flow data detection is an essential work.

Traditional data monitoring method has regression forecast method and the Kalman filter, etc. [1,2].These methods assume that the process is stable, the system is linear, the interference of the system is white noise, thus can obtain satisfactory result in linear system steady random time series predictions. However, the problem of urban road traffic flow is someone involved in the active system, has the characteristics of nonlinear and strong disturbance. Therefore, a nonlinear system model is more in line with the actual situation. Least squares abnormal data detection suitable for normal distribution characteristics of abnormal data, but is not suitable for direct application in the abnormal data of traffic flow detection. Therefore, directly using the least-square method, there will be a large amount of data was

wrongfully convicted. The current detection method of suitable for the unstable signal is wavelet analysis, this article uses the wavelet analysis method combined with least square method, make the ability to detect abnormal traffic flow data get great improvement.

2 The principle of wavelet denoising

2.1 The wavelet transform

The meaning of the wavelet transform is: call a basic wavelet function $\psi(t)$ to do after displacement τ, again under different scales a and the inner product of signal $x(t)$ to be analysed:

$$WT_x(A,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-\tau}{a})dt \quad a > 0 \qquad (1)$$

The equivalent frequency domain of formula (1) can be said such as:

$$WT_x(a,\tau) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{+\infty} X(\omega)\psi^*(a\omega)e^{j\omega\tau}d\omega \qquad (2)$$

In formula (2), $X(\omega)$ and $\psi(\omega)$ are respectively $x(t)$ and $\psi(t)$ Fourier transform.

Has the characteristics of multi-scale wavelet transform, the signal can be observed from coarse to fine gradually, and regard it as with basic frequency characteristic $\psi(\omega)$ for bandpass filter for signal filtering under different scales a.The choice of appropriate basic wavelet, making $\psi(t)$ in time domain is a limited support, $\psi(\omega)$ are more concentrated in the frequency domain, can make the wavelet transform in time domain and frequency domain has the ability of local signal

[a] Xiao Qian: xiaoqian_neu@163.com

characteristics, which is beneficial to detect signal singularity or eliminate noise.

2.2 Wavelet decomposition and reconstruction

In 1986, S.Mallat and Y.Meyer put forward the concept of multi-resolution analysis, and give the fast algorithm of wavelet decomposition and reconstruction, namely the Mallet algorithm.According to this algorithm,if f_k is the discrete sampling data of signal $f(t)$, $f_k = c_{0,k}$, the signal $f(t)$ decomposition formula for orthogonal wavelet transform is:

$$\begin{cases} c_{j,k} = \sum_n c_{j-1,n} h_{n-2k} \\ d_{j,k} = \sum d_{j-1,n} g_{n-2k} \end{cases} \quad (3)$$

Among them, $c_{j,k}$ ----scale coefficient;

$d_{j,k}$ ----wavelet coefficients;

h, g ---- a pair of quadrature mirror filter group (QMF);

j ----decomposition layers;

N ----discrete sampling point

Wavelet reconstruction process is the inverse operation of decomposition, the corresponding reconstruction formula is:

$$c_{j-1,n} = \sum_n c_{j,n} h_{k-2n} + \sum_n d_{j,n} g_{k-2n} \quad (4)$$

Wavelet multi-resolution analysis feature to signal under the different scales of multiresolution decomposition, and intertwined in different frequency of mixed signal is decomposed into different frequency band signal, so the signal has the ability of press band processing.

3 Least-square method based on wavelet analysis

Least squares method is a kind of mathematical optimization techniques, its data by minimizing the error sum of squares to find the best match function.Can be easily obtained by means of least square method of the unknown data, and make the the sum of the squares of the error between actual data and calculated data to a minimum.Least squares calculation process is as follows:

(1) Calculate the updated sequence of arithmetic average:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (5)$$

(2) Calculate each value of the residual error:

$$v_i = x_i - \bar{x} \quad (6)$$

(3) Using a sequence of standard error:

$$\sigma = \sqrt{\sum_{i=1}^{n} v_i^2 \Big/ (n-1)} \quad (7)$$

(4) According to the following formula, use La Yida of criteria to find abnormal data:

$$|v_i|_{\max} \le 3\sigma \quad (8)$$

Using the least square method can directly analyze the abnormal traffic flow data.Using the least square method, finding new sequence of arithmetic mean and then the residual error of individual values, using the residual error sequence of standard deviation, finally judge La Yida criterion and find the abnormal data. Least squares method is suitable for normal distribution characteristics of abnormal data detection, but is not suitable for direct application in the abnormal data of traffic flow detection. In this paper, firstly do wavelet analysis for traffic abnormal data to reconstruct the low frequency signal data, and then analyze the difference between the reconstructed sequence sequence x by the least squares method. Find abnormal data appear in the new position in the sequence, the sequence of abnormal data traffic at the corresponding position.

4 Abnormal traffic flow data simulation

4.1 Traffic data simulation based on least square method

In this paper, the simulation model of chongqing easyblog digital technology co., LTD. Production of GJK - 7 b a group of traffic flow data from the vehicle detector, each signal cycle record at a time (flow units: vehicle/cycle). In this paper, the simulation out a set of simulation data, and close to real data recorded 450 set of data, and according to the actual situation of artificial added four abnormal data points: 110 212.247.414.Using the least squares method for traffic signal data processing, be able to get the residual error, traffic signal and the residual error is shown in figure 1.
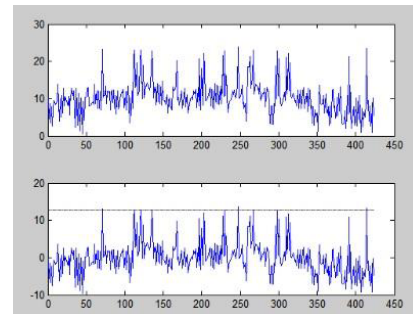

Figure 1 The traffic signal and residual error

4.2 traffic data simulation based on wavelet analysis

Selection based on the least square method of wavelet analysis to analyze traffic signal, choose db3 wavelet as the wavelet basis function, the traffic signal 2 layers wavelet decomposition scale, and then after wavelet decomposition of the low frequency part of the refactoring, the original traffic signal and reconstructing low frequency signal is shown in figure 2:
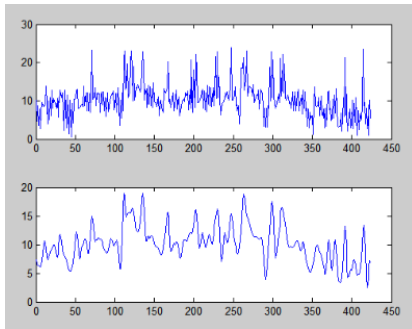
Figure 2 The traffic signal and reconstructing low frequency part

The refactoring least squares of low frequency part, reconstruct the difference sequences and residual error as shown in figure 3:
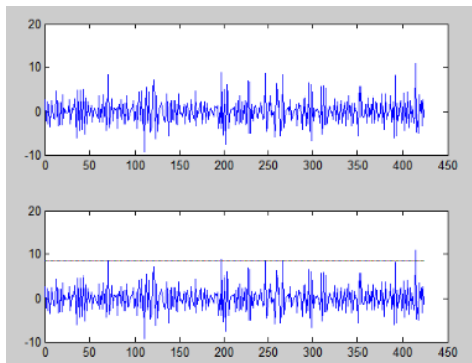

Figure 3 The difference sequences and residual error

4.3 the result analysis

In this paper, in the original traffic data added four abnormal data is as follows: 110 212.247.414, from figure 1 shows, using the least square method of traffic flow data detection, can find the abnormal data for 71.113.247.267.414, 247 and 414 of them are correct abnormal data points, and 71.113.267 is misjudgment, 110.212 is false negatives. Figure 3 shows, using wavelet analysis to detect the amount of traffic flow, you can find the abnormal data for 110.197.247.414, including 110.247.414 is correct, 197 is misjudgment, 212 is false negatives.

Table 1 abnormal data detection results contrast

The least square method	Least squares method based on wavelet analysis
Misjudgment rate of 60%	Misjudgment rate of 25%
False negative rate of 40%	False negative rate of 25%

The table 1 shows that the abnormal data by the least-square method to monitor the misjudgment rate was 60%, false negative rate was 40%. The least-square method based on wavelet analysis misjudgment rate was 25%, the false negative rate was 25%, higher than the least square method of accurate performance, so it can be seen that the least-square method based on wavelet

analysis of abnormal traffic flow data detection more accurate and effective.

5 Conclusion

In order to effectively manage the traffic planning and traffic control, need to be accurate and timely collection and traffic flow data, and detect the abnormal data.This article uses wavelet analysis combined with the traditional least squares method, so as to find abnormal point in the reconstructed signal data, for detecting abnormal data of non-stationary signal.Based on wavelet analysis and least square method to simulation experiment, the results show that the wavelet analysis has good precision.

References

1. Lei Shaolan, Sun Caixin, Zhou Quan, Zhang Xiaoxing. The Reserch of Local Linear Model of Short-Term Electrical Load on Multivariate Time Series[J]. Proceedings of the CSEE. 2006.26(2):25-28.
2. Zhang Jiangshan, Zhu Guangxi. Kalman Filter for Video Object Segmentation and Tracking[J]. Journal of Image and Graphics. 2002.7(6): 606-609.
3. Li Zhimin, Yi Liangyou, Xue Ping, Dai Gao, Xie Xianbao. Short-term traffic flow detection based on wavelet[J]. Application Research of Computers, 2011.18(5):1677-1678
4. Xiao ZhiXin，Yang YueXiang，Yang Lin. Anomaly Detection and Implementation Based on Wavelet[J]. Computer Science:2006.33(10):116-118
5. Xiong Wei. Synergetic phase transition detection method for network traffic anomolies based on wavelet[J]. Journal of Computer Applications. 2012.32(8):2271-2274
6. Song Yanbo, Peng Xiaoqi. Approach Based on Wavelet Analysis for Detecting and Amending Anomalous Samples in Data Set[J]. Mini-micro Systems.2011.2：325-329