JECET; December 2020-February 2021; Sec. C; Vol.10. No.1, 036-044 E-ISSN: 2278–179X
[DOI: 10.24214/jecet.C.10.1.03644]

Journal of Environmental Science, Computer Science and Engineering & Technology

An International Peer Review E-3 Journal of Sciences and Technology

Available online at www.jecet.org

Section C: Engineering & Technology

Research Article

Traffic change point detection and analysis

Jianmei Li[1], Chenchen Li[2], Xiaolin Wang[2], Ziwen Song[2], Rongji Zhang[2]

*School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo Shandong, China.

Received: 09 January 2021; **Revised:** 22 January 20210; **Accepted:** 30 January 2021

Abstract: Considering the time-varying and complexity of the actual traffic flow, this paper firstly introduces the theoretical knowledge of the change point analysis from the perspective of non-parametric methods in the case of unknown distribution. Then the related change point detection method is introduced. Finally, the dynamic algorithm is used to detect and estimate the actual number and position of the actual speed data of Beijing Road in Zhangdian District of Zibo City, and a result of troubleshooting the error change point is proposed. Finally, analyzed the reasons for the change point.

Keywords: Change point theory; traffic flow; traffic detection; trend; time series

CHANGE POINT THEORY

Definition and classification of change points: Change point theory is a classic branch of statistics. Its basic definition is that in a sequence or process, when a certain statistical characteristic (distribution type, distribution parameter) changes at a certain point in time by systemic factors rather than accidental factors. We call that point in time the change point. Change point identification is to use statistics or statistical methods to estimate the position of the change point [1, 2].

Generally speaking, the change point is "the point where one or some of the quantities in the model changes suddenly". Common change point models include mean change points (a sudden change in the mean value) and regression change points (a sudden change in the regression coefficient) , Probability change point (the probability of an event changes suddenly), etc. According to the detection method, the change point problem can be divided into parametric (Parametric) and non-parametric (Non-Parametric) change point detection; classified by the detection object, the change point problem can be It is divided into the detection of parameter change points in the distribution model and the change point detection of the distribution function itself; according to the data classification used, the change point problem can be divided into online detection (online data: Online Data) and offline detection (Offline Data: Off line Data).[3-6] According to the number of detection change points, the change point problems can be divided into single change point detection problems and multiple change point detection problems. Time series data sample values are arranged in the order of their appearance time, among which the change point of is that at an unknown moment, the moment when the distribution of the sample or other digital characteristics suddenly changes is the change point, and this moment is often unknown, so the main purpose of the research on the change point in the time series. It is to estimate the number and location of the change point, and at the same time often analyze the consistency of the change point estimator, the jump degree of the change point, the convergence speed, etc. The following formula is the condition that the data obey the change point in the Gaussian distribution.

$$Z_t = \begin{cases} \mu_1 & if\, 1 \le t \le \tau_1 \\ \mu_2 & if\, \tau_1 \le t \le \tau_2 \\ \dots & \dots \\ \mu_{k+1} & if\, \tau_k \le t \le \tau_{k+1} \end{cases}$$

Suppose there is a data set, and each data observation value is independent of each other. If at a certain moment, one or some variables in the model suddenly change, that is, there is a time point before which the data set conforms to a distribution after this point, the data set conforms to another distribution, then this point is the change point of the data set[7,8].

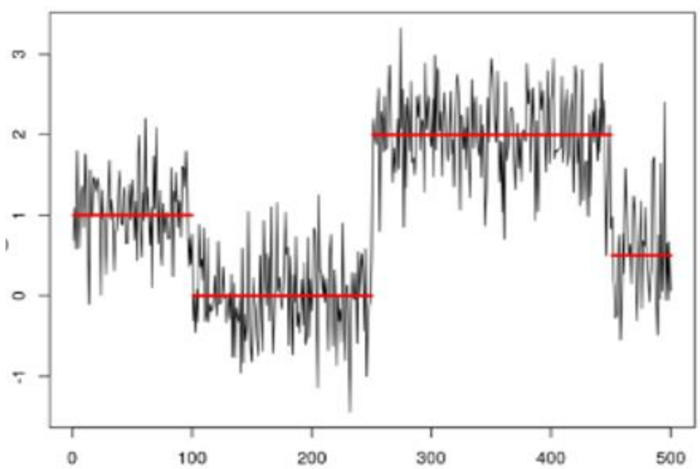

Figure 1: Change point diagram

Summary of data anomaly types

1. **Outlier:** Given the input time series x, the outliers are timestamp value pairs, where the observed value is different from the expected value of the time series.

DOI: 10.24214/jecet.C.10.1.03644

2. **Change Point:** Given the input time series x, the fluctuation point means that at a certain time "t", its state shows different characteristics from the values before and after "t" on this time series.
3. **Breakout:** The value at a certain moment in the sequential system is much steeper than the value at the previous moment.
4. **Anomalous:** Time Series Given a set of time series, the abnormal time series is the part of X that is inconsistent with most time series values.

MATERIALS AND METHODS

Change Point Detection: Change point detection is generally based on changes in the statistical characteristics of the signal sequence. Common statistical characteristics include the mean and variance of the previous data, residuals, etc. Obviously, some statistical characteristics will be changed when the signal sequence has a mutation. The more intuitive thing is that the mean and variance will change. The simplest way is that we can set the threshold reasonably based on this change, and we can determine when the threshold is exceeded the mutation occurs[9,10].

Classification of change point detection. From the point of view of monitoring purposes, it can be divided into mid-event change points (continuous sampling) and post-event change points (fixed samples). The former refers to continuously observing a certain random process, stopping the inspection when the change point is monitored, and not using future data, mainly used for event warning, the latter detecting the position of the past change point from the obtained time series sequence, mainly used for historical inspection . From the monitoring content, it is divided into single change point research and multiple change point research. Most of the current research focuses on the single change point area, that is, it is assumed that there is at most one change point in the time series data studied. However, the actual situation generally has multiple change points, which gives rise to the research problem of multiple change points. However, in the research stage, most scholars first determine the number of change points and then explore the location of the change points, which greatly reduces the research method[11].

General search methods for estimating and detecting change points: cumulative sum of squares method, least square method, and iterative cumulative sum of squares method, Bayes method, maximum likelihood method, local comparison method, wavelet analysis method, etc[12-13].

Change point detection method

1. **Control chart method CUSUM :** CUSUM detects whether the data distribution has changed based on the slight deviation of the accumulated data. This is also the oldest and most primitive, and is widely used in industrial quality inspection, automatic monitoring, and finance in the industry. For example, the monitoring solution used by Ali for the panoramic business platform.
2. **Probability Density Estimation：** Probability Density Estimation is for a set of time series, the probability density distribution before the change point and after the change point appears will be different.
3. **Direct Compute：** Because the probability density distribution is difficult to be accurate, it is derived that the probability density distribution is not calculated and evaluated, but the difference between the probability target distribution before and after a point. For the data before and after a

DOI: 10.24214/jecet.C.10.1.03644

point, some models and algorithms are used You can measure the difference in distribution before and after him. The difference of comparison is generally mean, variance, etc.

4. **Probability Method**: The Probability Method mainly compares the distribution of change points before and after. This part focuses on directly predicting whether a point is a change point. The methods used generally include Gaussian Process and Bayesian methods.
5. **Clustering Method**: Clustering Method regards the change point as a data set, which is divided into many categories according to the time series.

RESEARCH ON TRAFFIC FLOW CHANGE POINT

Traffic flow point: The traffic flow change point is "the point where one or some parameters in the traffic flow model change suddenly". This change is often connected with the abnormal state of the traffic flow, reflecting a certain qualitative change. Such as vehicle breakdowns, traffic accidents and other traffic incidents caused by traffic merges or traffic jams[14-16].

Using the change point analysis method, only the on-site data of traffic flow observations can be extracted to make a judgment on the existence of a sudden change, and according to a certain algorithm, the number of sudden changes and the specific location of the sudden change (spatial location or time location) can be calculated. Point estimation or interval estimation, etc., so as to judge whether there is abnormal traffic flow, and implement timely and reasonable guidance through advanced information dissemination technology. For various traffic flow parameters, such as flow, speed, lane occupancy, etc., when using this method, there is no special requirement for its distribution, and considering that the result of sudden change often leads to changes in the distribution of traffic flow parameters and other factors.

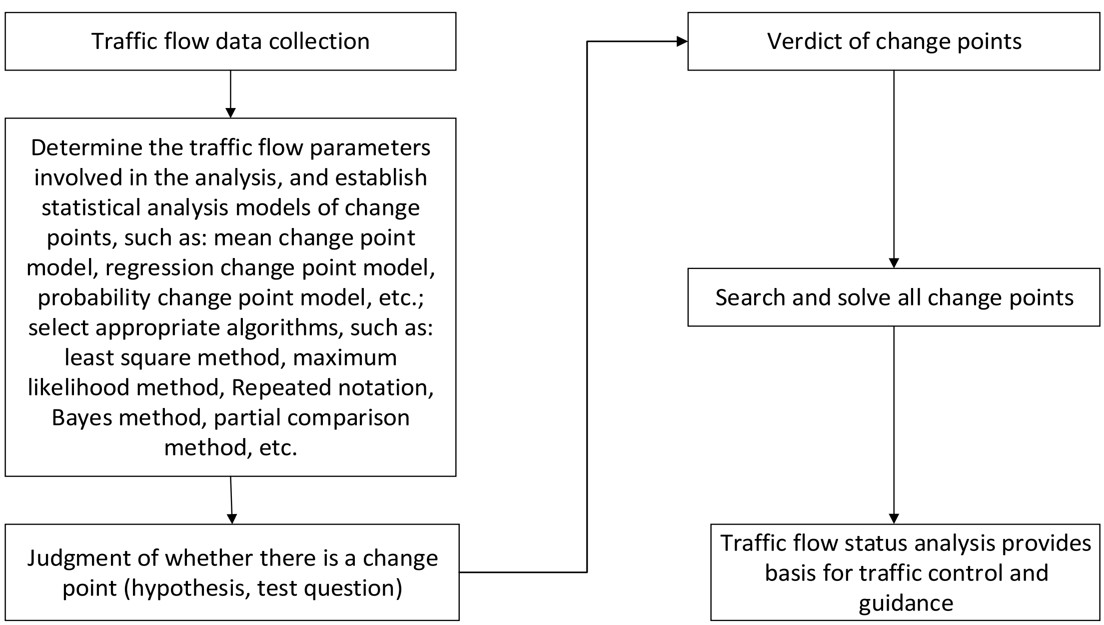

Figure 2: Analysis steps of traffic flow point

JECET; December 2020-February 2021; Sec. C; Vol.10. No.1, 036-044.
DOI: 10.24214/jecet.C.10.1.03644

Change point detection of traffic flow: This article introduces a selective investigation of algorithms for offline detection of multiple change points in a multivariate time series, using a common structured method strategy to organize this huge work.

More precisely, the detection algorithm considered in this review has three elements: cost function, search method and constraints on the number of changes. Each of these elements is individually described, reviewed and discussed. The main algorithm principles described in this article are provided in the R package named Man pages for trend in CRAN (Comprehensive R Archive Network).

The following mainly introduces several algorithms related to change point detection. MK inspection is mainly trend detection. Pettitt, Buishand, and SNHT are time series homogeneity tests. Homogeneity testing allows you to determine whether a series can be considered homogeneous over time, or whether there is a time for change.

1. **MK inspection:** The purpose of the Mann-Kendall (MK) test is to statistically evaluate whether the variable we are interested in has a monotonous upward or downward trend over time. The MK test does not require the data to be normally distributed, nor does it require that the trend of change—if it exists—is linear. If there are missing values or values below one or more detection limits, MK detection can be calculated, but the detection performance will be adversely affected. The independence assumption requires that the time between samples is large enough so that there is no correlation between the measured values collected at different times.
2. **Standard Normal Homogeneity Test (SNHT):** The SNHT test (standard normal homogeneity test) was developed by Alexandersson (1986) to detect changes in a series of rainfall data. The test is applied to a series of ratios, comparing the observations at the measuring station with the average of several stations. Then standardize the ratio. The Xi series here corresponds to a standardized ratio. (Use Monte Carlo resampling to provide p-value).
3. **Pettitt Test**：Use Pettitt Test for non-parametric test to test the changes in the central tendency of the time series. Pettit's change point detection test statistic is the maximum value of the absolute value of the vector, and the approximate probability of the two-sided test is calculated. In general, $P \leq 0.5$.
4. **Buishand U Test**：Buishand U Test test change point detection. The data is a normal random variable, and then a model with a single shift (change point) can be proposed, which mainly uses the mean and sample standard deviation to detect.

Program realization of detection method: Because python has a huge open source library, python also inherits many functions of the R language package. The following mainly uses the above-mentioned Mann-Kendall (MK), Standard Normal Homogeneity Test (SNHT), Pettitt Test, Buishand U on the basis of python.

Test four change-point detection methods to detect and compare the change-points on the evening peak speed data of Beijing Road in Zibo City, and analyze the detection results. See the appendix for the specific program code. The source data is shown below:

DOI: 10.24214/jecet.C.10.1.03644

通过道口时间	经过时间	通过交叉口时间	速度
20190101073433	20190101073140	173	31
20190101073433	20190101073146	167	32
20190101073435	20190101073134	181	30
20190101073436	20190101073118	198	27
20190101073437	20190101073102	215	25
20190101073438	20190101073127	191	28
20190101073439	20190101073153	166	33
20190101073440	20190101073152	168	32
20190101073441	20190101073219	142	38
20190101073441	20190101073148	173	31
20190101073442	20190101073220	142	38
20190101073444	20190101073151	173	31
20190101073446	20190101073045	241	22
20190101073448	20190101073157	171	32
20190101073647	20190101073411	156	35
20190101073648	20190101073405	163	33
20190101073652	20190101073419	153	35
20190101073655	20190101073353	182	30
20190101073701	20190101073357	184	29

Figure 3: Traffic source data of Beijing Road in Zibo City

Import the above speed into the python program and display it visually.

```
"E:\anaconda 1\python.exe" "E:/python 文件/变点分析3.py"
Mann-Kendall: [59, 92, 99]
Pettitt: 37
Buishand U Test: 37
Standard Normal Homogeneity Test (SNHT): 14


Process finished with exit code 0
```

Figure 4: The results of the program

DOI: 10.24214/jecet.C.10.1.03644

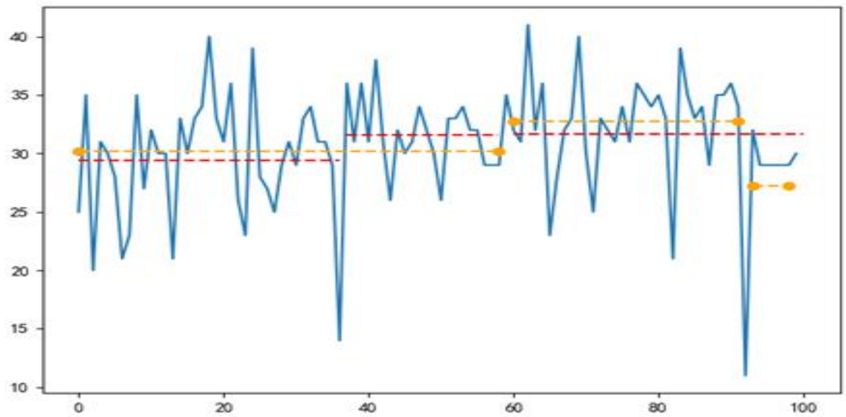

Figure 5: Speed change graph

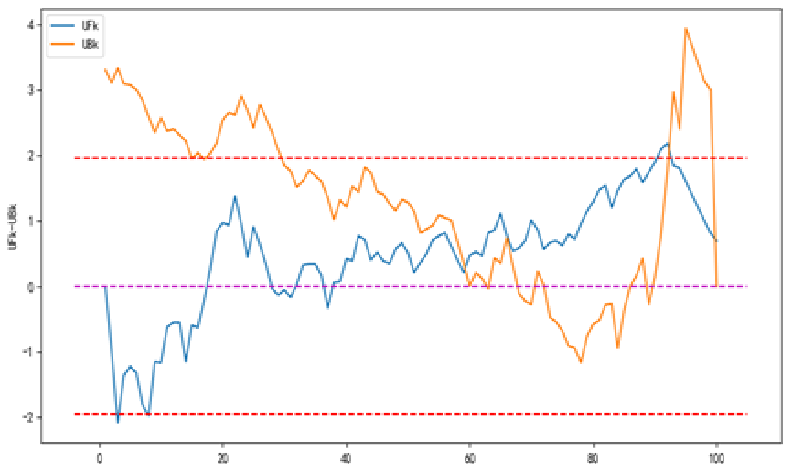

Figure 6: Kendall_change_point_detection positive and negative sequence cross graph

From the above, the obtained change points are 14, 37, 59, 92, 99. However, it can be seen from the change trend of the graph that the position of 92 is too much affected by the abnormal point, and the position of 99 is on the edge, so the results of the change points at 92 and 99 are not highly reliable. The position of the change point of 14, 37, 59 is more accurate based on actual experience. Generally, it is affected by the control after traffic absorption or the influence of individual vehicles, which leads to obvious changes in subsequent vehicle speeds.

CONCLUSION

This article introduces the change-point theory and change-point detection theory that are often used in data analysis, and detects and analyzes the changes in the data combined with related traffic flow data.

First of all, when using the detection algorithm, as shown in the above example, abnormal points and jump points have a huge impact on the detection of change points, so the change of jump degree during the mutation inspection process should be considered. When the test result shows that the change point should exist, but according to the change of the trend chart and actual experience, the possibility of its occurrence is very small, and the corresponding location point should be excluded. In addition, the detection and clustering methods of change points have the same effect in the extraction of the continuous data law of time series. The clustering method can be further used to verify the detection results.

Finally, comprehensive use of multiple traffic flow parameters, multiple detection algorithms, and strengthen the comprehensive utilization of data from various stations on the road network, increase the analysis of data from multiple perspectives such as time and space (if conditions permit). It is an effective way to improve the level of traffic flow mutation analysis.

REFERENCES

1. Marcel Westerman, Remco Litjens, Jean-Paul Linnartz. Integration of Probe Vehicle and Induction Loop Data-Estimation of Travel Times andAutomatic Incident Detection[R]. California PATH Research Report, UCB-ITS-PRR-96-13, 1996.
2. Baher Abdulhai, Stephen Ritchie, Mahadevan Iyer. Implementation of Advanced Techniques for Automated Freeway Incident Detection[R].California PATH Research Report, UCB-ITS-PRR-99-42, 1999.
3. B.Abduhai, S.Ritchie, Enhancing the universality and transferability of freeway incident detection using a bayesian based neural network [J].Transportation Research Part C, 1999, 36(6): 261-280.
4. I.E.Auger, C.E.Lawrence, Algorithms for the optimal identification of segment neighborhoods [J]. Bulletin of mathematical biology, 1989, 51(1): 39-54.
5. B.S.Darkhovshk, A non-parametric method for the a posteriori detection of the disorder time of a sequence of independent random variables ［J］． Theory of Probability and Its Applications，1976，21: 178-183．
6. M. Csörgő, L. Horvath L. 20 Nonparametric methods for changepoint problems [J]. Handbook of statistics, 1988, 7: 403-425.
7. D.L. Hawkins, Retrospective and sequential tests for a change in distribution based on kolmogorov-smirnov-type statistics ［J］ Sequential Analysis: Design Methods and Applications，1988，7: 23-51
8. P.R.Krishnaiah, B.Q.Miao, Review about estimates of change-point[J].Handbook of Statistics,1988,7:375-402
9. G.M.Chen, Y.K.Choi, Y. Zhou, Nonparametric estimation of structural change points in volatility models for times series[J] Jouanal of econometrics,2005,126:79-114
10. J.V.Braun, R.K.Braun, H.G.Muller, Multiple change point fitting via quasilikelihood with application to DNA sequence segmentation [J.Biometrika, 2000, 87(2):301-314.
11. S. Cheon, J. Kim, Multiple change-point detection of multivariate mean vectors with the Bayesianapproach [J]. Computational Statistics and Data Analysis, 2010, 54(2): 406-415.
12. Chen Xiru, Inference in a sample change point model][J].Scientia Sinica ,(Series A),1988, 6:654-667.

DOI: 10.24214/jecet.C.10.1.03644

13. TAN Chang chun, MIAO Bai qi. Inference in T-distribution with at most one change point []J].Journal of University of Science and Technology of China,2005、35(1):51-58.
14. K. Frick, A. Munk and H. Sieling, Multiscale change-point inference [J].J. R. Statist. Soc. B, with discussion and rejoinder by the authors, 2014, 76(3): 495-580.
15. A. Li H Munk , H. Sieling , FDR-Control in Multiscale Change-point Segmentation[J]. Statistics, 2016.10:918-959.
16. R. Beran, Local Extremes, Runs, Strings and Multiresolution: Discussion [J].Annals of Statistics, 2001, 29(1):99-107.

Corresponding author: Jianmei Li;

Postgraduates, School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo Shandong, China

Online date of publication: 30.01.2021

DOI: 10.24214/jecet.C.10.1.03644