The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

Comparison of Change Detection Methods for Ongoing Time Series Data
---- Extended SPRT, Chow Test, Extended DP ----

*Hiromichi Kawano, **Tetsuo Hattori, **Katsunori Takeda, ***Tetsuya Izumi

(* NTT Advanced Technology/ Musashino-shi Nakamachi 19-18, Tokyo, Japan

**Kagawa University/ 2217-20 Hayashi-Cho, Takamatsu City, Kagawa, Japan

*** Micro Technica Co., Ltd./ Yamagami BLD. 3-12-2 Higashi-ikebukuro, Toshima-ku, Japan)

(hattori@eng.kagawa-u.ac.jp)

Abstract: For change point detection of time series data, we have already proposed an application of Sequential Probabilistic Ratio Test (SPRT). In addition, we have proposed a Dynamic Programming (DP) method as well, for the case where we have not only to detect the change point, but also to take into account an action cost after the detection. In this paper, we show the effectiveness and differences of the two extended methods (ESPRT and EDP) in comparison with the well-known Chow Test, by experimental results.

Keywords: Time series analysis, Dynamic Programming, sequential probability ratio test (SPRT), Chow Test

I. INTRODUCTION

Change point detection (CPD) problem in time series is to identify whether the generation structure of monitoring data has changed at some time point by some reason, or not. We consider that the problem is very important and that it can be applied to a wide range of application fields. [1].

We have already proposed an application of Sequential Probabilistic Ratio Test (SPRT) and a Dynamic Programming (DP) method for the case where we have not only to detect the change point, but also to take into account an action cost after the detection. And we have presented the effectiveness of the two methods in comparison with the well-known Chow Test by using multiple regression model [2]. We extend the definition of the structural change point in the SPRT method (ESPRT), and show the improvement of the change detection accuracy [2]. In this paper, we extend the definition of the structural change point in the DP method (EDP) and we show the effectiveness and differences of the two extended methods (ESPRT and EDP) in comparison with the well-known Chow Test, by experimental results.

II. DP method, SPRT and Chow Test

1. DP method and SPRT

For simplicity, we explain the detection methods using a linear single regression model as shown in Fig.1.

2. DP method model ([3])

The concrete procedure of structural change detection is as follows (see an example of time series data in Fig.1).

Step1: Make a prediction expression and set the tolerance band (a) (e.g. a=2σs) that means permissible error margin between the predicted data and the observed one.

Step2 : While monitoring the coming data, if the data comes into a specified tolerance zone, then we call the situation "in", or "hitting", otherwise "out" or "failing". Based on this monitoring, we can judge that the structure of the time series data has changed, if the failing occurs by continuing N times. This specified tolerance is defined as, e.g., 2σ of the distribution error as shown in the Fig.1.

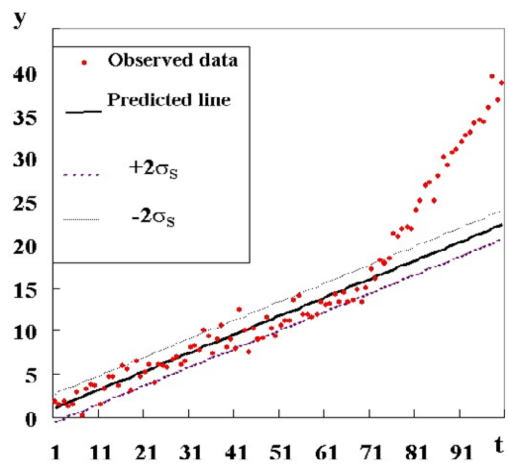

Fig.1. Example of time series data where the change point tc* =70.

3. Cost function and optimal solution

Let the cost (n) be $a \cdot n$ an as a linear function for *n*, where "*a*" is the loss caused by the failing in one time. And for simplicity, let *T* and *A* denote the Total cost and Action cost. The Action cost means the cost when some action has been taken at the time of structural change detection. Then, the evaluation function is denoted as follows.

$$T = A + a \cdot n \qquad (1)$$

We recursively define a function $ET(n, N)$ to obtain the optimum number of times *n* that minimizes the expectation value of the evaluation function of T, using the concept of DP. Let *N* be the optimum number. Let the function $ET(n, N)$ be the expectation value of the Total cost at the time when the failing has occurred in continuing *n* times, where *n* is less than or equal to *N*, i.e., $0 \leq n \leq N$.Then, the function is recursively given as in the following.

$$(n = N) \;\; ET(n, N) = A + a \cdot N \qquad (2)$$

$$(n < N) \;\; ET(n,N) = P(\overline{S}_{n+1} \mid S^n) \cdot a \cdot n + (1 - P(\overline{S}_{n+1} \mid S^n)) \cdot ET(n+1, N) \qquad (3)$$

where S^n is the state of failing in continuing *n* times, $\overline{S}_{n+1}$ is the state of hitting for the (*n*+1) th time observed data, and $P(\overline{S}_{n+1} \mid S^n)$ is the conditional probability that the state $\overline{S}_{n+1}$ occurs after the state S^n occurred.

For the aforementioned *ET(0,N)*, the following theorem holds, and gives the n that minimizes the expectation value of the evaluation function of Eq. (1).

Theorem ([3]).

The N that minimizes *ET(0,N)* is given as the largest number n that satisfies the following Inequality (4).

$$a < (A + a) \cdot P(\overline{S}_n \mid S^{n-1}) \qquad (4)$$

4. Procedure of SPRT ([2])

The concrete procedure of structural change detection is as follows (see an example of time series data in Fig.1).

Step2: Set up the null hypothesis H_0 and alternative hypothesis H_1. H_0 means change has not occurred yet. H_1 means Change has occurred. Set the values α, β and compute *C1 (= β/(1-α))*and *C2 (= (1-β)/α)*, Initialize i = 0, $\lambda_0 = 1$.

Step3: Incrementing i (i = i+1), observe the following data y_i. Evaluate the error $|\varepsilon_i|$ between the data y_i and the predicted value from the aforementioned prediction expression.

Step4: Judge as to whether the data y_i goes in the tolerance band or not, i.e., the ε_i is less than (or equal to) the permissible error margin or not. If it is Yes, then set $\lambda_i = 1$ and return to Step3. Otherwise, advance to Step5.

Step5: Calculate the probability ratio λ_i, using the following Equation (5) .

$$\lambda_i = \lambda_{i-1} \frac{P(\varepsilon_i \mid \mathbf{H}_1)}{P(\varepsilon_i \mid \mathbf{H}_0)} \qquad (5)$$

where, if the data y_i goes in the tolerance band, $P(\varepsilon_i | H_0) = \theta_0$ and $P(\varepsilon_i | H_1) = \theta_1$, otherwise, $P(\varepsilon_i | H_0) = (1-\theta_0)$ and $P(\varepsilon_i | H_1) = (1-\theta_1)$.

Step6: Execution of testing.

(i) If the ratio λ_i is greater than C_2, dismiss the null hypothesis H_0, and adopt the alternative hypothesis H_1, and then End.

(ii) Otherwise, if the ratio λ_i is less than C_1, adopt the null hypothesis H_0, and dismiss the alternative hypothesis H_1, and then set $\lambda_i = 1$ and return to Step3.

(iii) Otherwise (in the case where $C_1 \leq \lambda_i \leq C_2$), advance to Step7.

Step7: Observe the following data y_i incrementing i. Evaluate the error $|\varepsilon_i|$ and judge whether the data y_i goes in the tolerance band, or not. Then, return to Step5 (calculation of the ratio λ_i).

5.Chow Test ([2])

The well known Chow Test checks the significant differences among residuals of three Regression Lines, where regression Line 1 obtained from the data before a change point tc, Line 2 from the data after tc, and Line 3 from the whole data so far, by setting up hypothesis of change point at each point in the whole data.

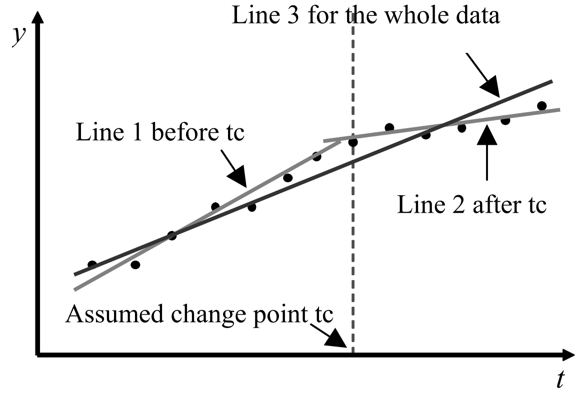

Fig.2. Conceptual image of Chow Test

III. Extended DP and SPRT

1. Extended DP

The DP detects a change point at the time when the expectation value of the evaluation function of Eq. (1) is minimized. Then, the detected change point equals to the terminated time point and its detection tends to be delayed from true change point. Thus in this section, we extend the definition of detected change point by DP method. As such extension, we adopt the number tc-M where tc is ordinary aforementioned change point and M is the number of times when the observed data continuously goes of tolerance zone until the n that minimize the expectation value. M is given by using the posterior probability $P(E_{cn}|S_n)$. According to the Bayes' theorem, the posterior probability $P(E_{cn}|S_n)$ is given by the following (6).

$$P(E_{cn}|S_n) = \frac{P(S_n \cap E_{cn})}{P(S_n \cap E_{cn}) \cup P(S_n \cap E^n)}$$

$$= \frac{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i}}{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i} + (1-\lambda)^n R^n}$$

$$= \frac{1}{1 + \frac{(1-\lambda)^n R^n}{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i}}} = \frac{1}{1 + D(n)} \qquad (6)$$

where $D(n) = \frac{(1-\lambda)^n R^n}{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i}}$

and λ means the probability of the structural change occurrence and R is the probability of the failing when the structure is unchanged, Rc is the probability of the failing when the structural change occurred.

2.Etended SPRT([2])

The SPRT detects a change point at the time when the probabilistic ratio λ_i is greater than C_2 (=(1-β)/α). Then we adopt the number tc-***M*** where tc is ordinary aforementioned change point and ***M*** is the number of times when the observed data continuously goes of tolerance zone until the ratio $\lambda_i > C_2$ The number ***M*** can be obtained from the equation (7).

$$\left(\frac{\theta_1}{\theta_0}\right)^M > C_2 \left(= \frac{1-\beta}{\alpha}\right) \qquad (7)$$

Then we have the equation (7) using Gauss notation. So, the value of ***M*** depends on the parameters (Table 2). That is, ***M***=2 (case **a)**, ***M***=3 (case **b)**, ***M***=4 (case **c)**.

$$M = \left\lceil \log_{\frac{\theta_1}{\theta_0}} \frac{1-\beta}{\alpha} \right\rceil$$

IV. Experimentation

In our experimentation, time series data is generated by the following equations.

$$y = a_{11}x_1 + a_{12}x_2 + b + \varepsilon \qquad (t \le t_c^*) \qquad (8)$$

$$y = a_{21}x_1 + a_{22}x_2 + b + \varepsilon \qquad (t_c^* \le t) \qquad (9)$$

where ε~N(0,σ^2), i.e., the error ε is subject to the Normal Distribution with the average 0 and the variation σ^2, and **t**c* means the change point. In addition, we have set **t**c*=70.

Table 1 Equations for generating time series

Data No.	equation (time t=1,2,…,69)	equation (time t=70,71,…,100)	σ
1	$y = 2x_1 + 3x_2 + 10$	$y = 3x_1 + 3x_2 + 10$	5
2		$y = 2.5x_1 + 3x_2 + 10$	5
3		$y = 2.5x_1 + 3x_2 + 10$	1

Table 2 Parameter values in SPRT

Case	Data No.	α	β	θ_0	θ_1
1-a	1	0.05	0.05	0.1	0.9
1-b				0.2	0.8
1-c				0.3	0.7
2-a	2	0.05	0.05	0.1	0.9
2-b				0.2	0.8
2-c				0.3	0.7
3-a	3	0.05	0.05	0.1	0.9
3-b				0.2	0.8
3-c				0.3	0.7

We have experimented with DP method and Chow Test for the artificial data based on the above equations (8) and (9).

1. Experimental conditions

(i) Tolerant error: ±2σ of the distribution on error ε.
(ii) The concrete values of parameters are shown in Table 1.(Fig.4 shows an example of the graph of generated time series data by the above equations.).
(iii) Repetition times for making sets of data: M=100.
(iv) Parameters setting for detection:
(a) SPRT: parameters are shown in Table 2
(b) DP method: λ =0.01, A/*a* = 10, 20.
(c) Chow Test: **s**ignificance level (α=0.05) for testing.

Table 3. Change detection point.

Data No.	method	SPRT			DP		ChowTest
	conditions	θ_0=0.1, θ 1=0.9	θ_0=0.2, θ 1=0.8	θ_0=0.3, θ 1=0.7	A/a=10	A/a=20	α=0.05
1	Average	69.25	74.50	76.54	68.95	74.14	76.50
	Standard deviation	6.35	4.73	4.97	6.54	4.04	16.02
2	Average	75.39	83.84	86.88	74.59	83.30	76.60
	Standard deviation	11.70	9.22	7.98	11.51	8.71	16.16
3	Average	68.85	71.88	73.53	68.61	71.73	76.50
	Standard deviation	6.02	1.90	1.70	6.27	1.78	16.02

Table 4. Estimated change point.

Data No.	method	SPRT			DP		ChowTest
	conditions	θ_0=0.1, θ 1=0.9	θ_0=0.2, θ 1=0.8	θ_0=0.3, θ 1=0.7	A/a=10	A/a=20	α=0.05
1	Average	67.62	69.25	69.01	67.95	72.14	58.21
	Standard deviation	6.09	2.42	2.35	6.54	4.04	20.38
2	Average	73.39	77.93	77.35	73.71	81.24	57.09
	Standard deviation	21.24	9.42	8.46	11.42	8.72	20.95
3	Average	67.58	69.21	68.94	67.61	69.37	58.35
	Standard deviation	6.05	2.35	2.29	6.27	5.36	20.38

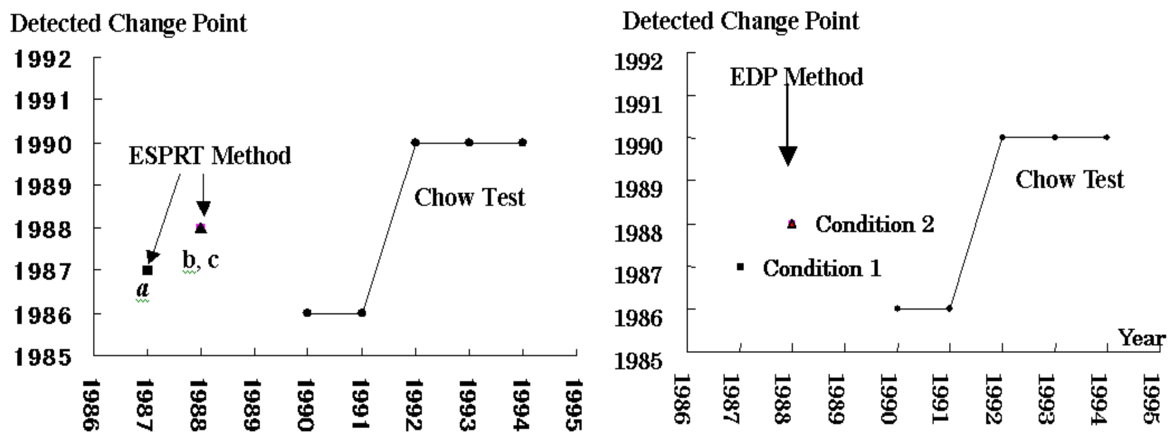

Fig.3.Results for Power supply time series.

2. Results for artificial data

The results are shown in Table 3(Change detection point) and Table 4(estimated change point using ESPRT and EDP). Those results are based on the average value and standard deviation of 100 times computation. It is expected that the change point will be detected around the time at t=70.

Applying the extended definition, we obtain the improvement of the change detection accuracy

3. Results for real data

Results for the time series data of power generation quantity in Japan are shown in Fig.3. In 1986 (Condition 1 denotes A/a=10, Condition 2 denotes A/a=20). The "oil shock" happened in the worldwide and oil prices suddenly rose. We can see that Chow Test correctly detects the change point of 1986, at 1990 and 1991. Until then, it detects no change point. But after 1991, it outputs 1990 as change point. Both of ESPRT and EDP can estimate change point.

IV. CONCLUSION

We have made a comparison between both of ESPRT and EDP method and Chow Test using multiple regression modeled time series data and real data. We have found that ESPRT and EDP method works very well. As further study, we investigate the relation between the parameters of ESPRT and EDP method.

REFERENCES

[1] J.Chen and A.K.Gupta (2000), Parametric Statistical Analysis, Birkhauser.

[2] Katsunori Takeda, Tetsuo Hattori, Izumi Tetsuya, Hiromichi Kawano (2010), Extended SPRT for Structural Change Detection of Time Series Based on Multiple Regression Mode, Proc. of AROB 15th '10, pp.755-758.

[3] Hiromichi Kawano, Tetsuo Hattori, Ken Nishimatsu (2004), "Structural Change Detection in Time Series Based on DP with Action Cost", Proc. of the 2004 IEEE Intern'l Conf. on IRI, pp.402-407.