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

DP method for Structural Change Detection as Optimal Stopping ---- Experimentation in Multiple Regression Model ---

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

(*Musashino-shi Nakamachi 19-18, Tokyo 180-0006, Japan

NTT Advanced Technology,

**2217-20 Hayashi-Cho, Takamatsu City, Kagawa 761-0396, Japan

Graduate School of Engineering, Kagawa University,

***Yamagami BLD. 3-12-2 Higashi-ikebukuro, Toshima-ku, Tokyo 170-0013, Japan

Micro Technica Co., Ltd.)

Abstract: If a model begins to fail to prediction of the time series, we have to detect such a structural change (i.e., disparity between the prediction model and the data) quickly and correctly, and to renew the prediction model after the change detection as soon as possible. In this paper, we formulate the structural change detection of time series as an optimal stopping problem, using Dynamic Programming (DP). Moreover, we present experimental results of the change point detection in multiple regression modeled time series data, comparing with SPRT and Chow Test.

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

I. INTRODUCTION

Structural change detection for ongoing time series data as early and correctly as possible is a very important problem in a practical sense, especially in the field of quality control and management of something that depends on time. For example, early degradation detection of the quality in communication system, manufactures from production lines in a factory, and etc., are such kind of serious problems [1].

We have already proposed an application of Sequential Probabilistic Ratio Test (SPRT). Moreover, we have proposed 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. Our experimentation has been done just by using single regression model [2]. However, multiple regression models are more generally used for time series data analysis than single regression one. And also, the theory of Chow Test is based on the general multiple linear regression itself.

In this paper, we examine how the DP method, SPRT and Chow Test work for the change detection of multiple regression model based data, by experimentation.

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. 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.

2.DP method model ([2])

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.

We assume that the structural change is Poisson occurrence of average, and that, once the change has occurred during the observing period, the structure does not go back to the previous one (Fig.2). The reason why we set such a model is that we focus on the detection of the first structural change in the sequential processing.

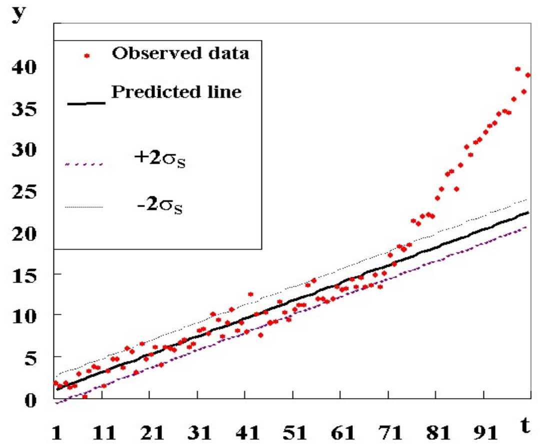
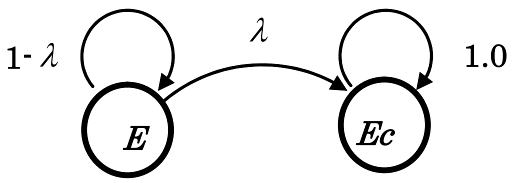
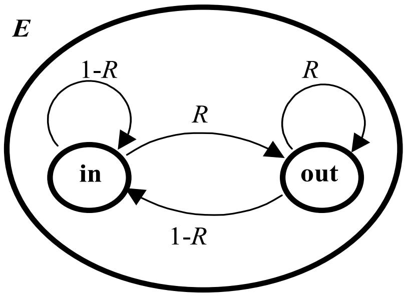


Fig.1. Example of time series data where the change poi nt tc* =70.

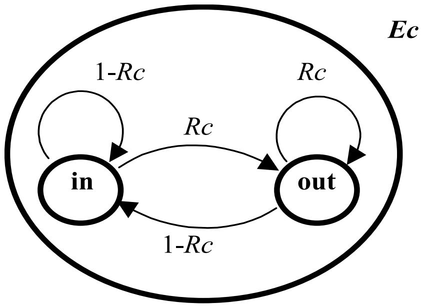


Ec : State that the structural change occurred.
E : State that the structure is unchanged.
λ : Probability of the structural change occurrence. **(Poisson Process.)**

(a) Structural change model



(b) Internal model of the State *E*.



(c) Internal model of the State *Ec*.

Fig.2. Structural change model of time series data.

3. Definition of cost function

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..

$$T = A + a \cdot n$$

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.

$$(n = N)\ ET(n, N) = A + a \cdot N \quad (1)$$

$$(n < N)\ ET(n,N) = P(\overline{S}_{n+1} \mid S^n) \cdot a \cdot n + (1 - P(\overline{S}_{n+1} \mid S^n)) \cdot ET(n+1, N) \quad (2)$$

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. Then, from the definition of the function $ET(n,N)$, the goal is to find the *N* that minimizes $ET(0,N)$, because the *N* is the same as *n* that minimizes the expectation value (Eq.(2)).

4. Optimal solution

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. (2).

Theorem (**[2]**).

The N that minimizes ET(0,N) is given as the largest nu mber n that satisfies the following Inequality (3).

$$a < (A + a) \cdot P(\overline{S}_n \mid S^{n-1}) \quad (3)$$

5. Procedure of SPRT ([3])

The concrete procedure of structural change detection is as follows (see 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 C_1 (= β/(1-α))and C_2 (= (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 (4) .

$$\lambda_i = \lambda_{i-1} \frac{P(\varepsilon_i \mid \mathbf{H}_1)}{P(\varepsilon_i \mid \mathbf{H}_0)} \quad (4)$$

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 $C_1 \le \lambda_i \le C_2$), 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).

6.Chow Test ([3])

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.

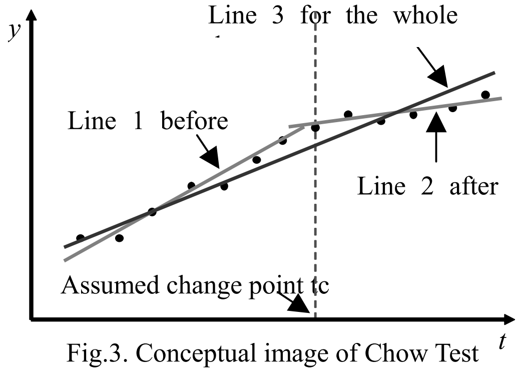


Fig.3. Conceptual image of Chow Test

III. Experimentation

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

$$y = a_{11}x_1 + a_{12}x_2 + b + \varepsilon \quad (t \le t_c^*) \quad (5)$$

$$y = a_{21}x_1 + a_{22}x_2 + b + \varepsilon \quad (t_c^* \le t) \quad (6)$$

where $\varepsilon \sim N(0,\sigma^2)$, i.e., the error ε is subject to the Normal Distribution with the average 0 and the variation σ^2, and **tc*** means the change point. In addition, we have set **tc***=70.

We have experimented with DP method and Chow Test for the artificial data based on the above equations (5) and (6).

1. Experimental conditions

(i) Tolerant error: $\pm 2\sigma$ 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: $\lambda = 0.01$, $A/a = 10, 20$.

(c) Chow Test: significance level ($\alpha = 0.05$) for testing.

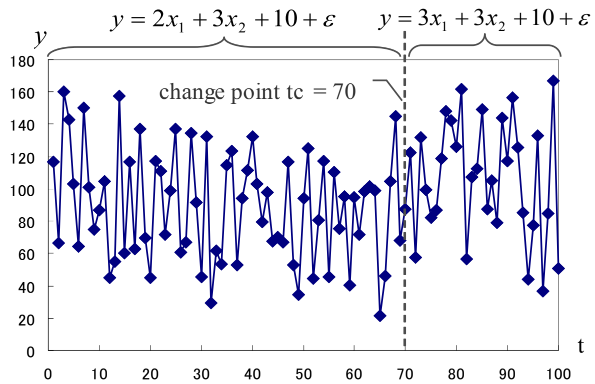


Fig.4. Example of time series data in Data 1.
(True change point: t=70.)

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

2. Results

The results are illustrated in Fig.5, where horizontal axis shows observation time t (observation is started after t=40) and vertical axis shows the time when the change point was detected. Those results are based on the average value of 100 times computation. It is expected that the change point will be detected around the time at t=70.

We have verified that the two methods (SPRT and OS method) meet our intuition very well as follows.

(i) The graph of Chow test takes continuous value for time t. This means that Chow test detects change at every time when data is observed.

(ii) From Fig.5-Fig.7, even after the enough time passes, Chow test cannot detect change point properly.

(iii) Both of SPRT and DP method can detect a change point more suitably than Chow Test.

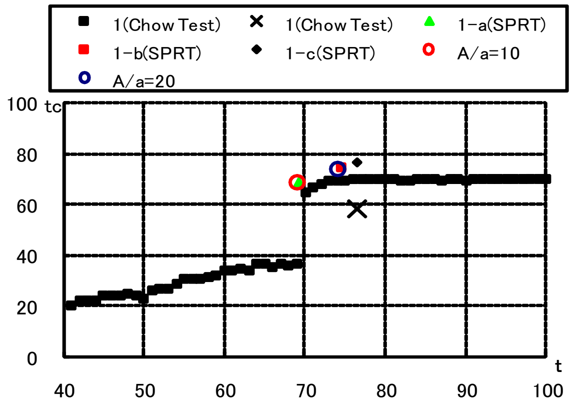
(iv) Both of SPRT and DP method have a tendency to detect a change point early when the σ is small. This is because, when the σ becomes smaller, the probability that exceeds the tolerant interval (2σ) becomes larger, i.e., the "failing" easily occurs.

IV. CONCLUSION

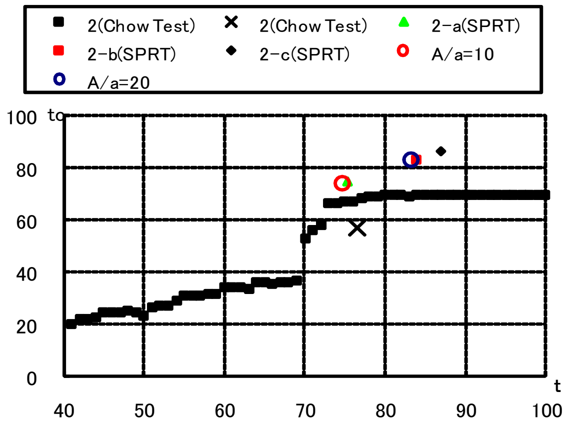
We have made a comparison between both of SPRT and DP method and Chow Test using multiple regression modeled time series data. We have found that SPRT and DP method can detect the change point very well for the real time ongoing data. Although the DP method depends on the ratio A/a (A: Action cost, a: loss cost), we can expect that it works well in a practical sense. As further study, we investigate the relation between the parameters of SPRT and OS method.

REFERENCES

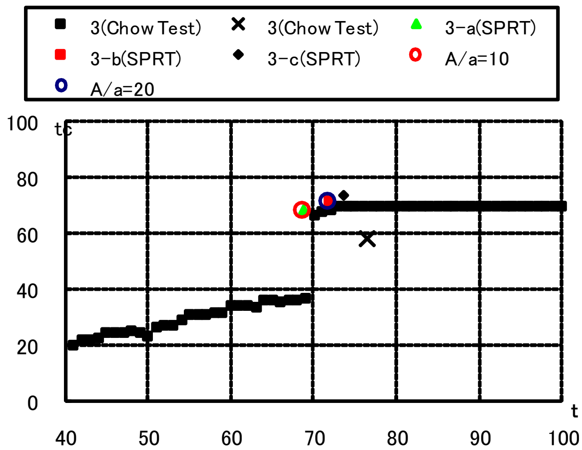
[1] J.Chen and A.K.Gupta (2000), Parametric Statistical Analysis, Birkhauser.
[2] Hiromichi Kawano, Tetsuo Hattori, Ken Nishimatsu (2004), "Structural Change Detection in Time Series Based on DP with Action Cost", Proc. of the 2004 IEEE IRI, pp.402-407.
[3] Katsunori Takeda, Tetsuo Hattori, Izumi Tetsuya, Hiromichi Kawano (2010), Extended SPRT for Structural Change Detection of Time Series Based on Multiple Regression Mode, Proc. of AROB15th'10, pp.755-758.



(1) Case1



(2) Case 2



(3) Case 3

Fig.5. Relation between the observing time t and detected change point tc.