Early Structural Change Detection as an Optimal Stopping Problem (I) --- Formulation Using Dynamic Programming with Action Cost ----

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

*Graduate School of Engineering, Kagawa University / 2217-20 Hayashi, Takamatsu City, Kagawa 761-0396, Japan
** Micro TechnicaCo., Ltd./ Yamagami BLD. 3-12-2 Higashi-ikebukuro, Toshima-ku, Tokyo 170-0013, Japan
*** NTT Advanced Technology / Musashino-shi Nakamachi 19-18, Tokyo 180-0006, Japan
(Tel : 81087-864-2221; Fax : 81-087-864-2262)
(hattori@eng.kagawa-u.ac.jp)

Abstract: Even if an appropriate prediction expression and/or model are constructed to fit a time series, the model gradually begins to fail the prediction of the time series from some time point. In such case, it will be important not only to quickly detect the failing situation but also to renew the prediction model after the detection as soon as possible. In this paper, we formulate the structural change detection in time series as an optimal stopping problem, using the concept of DP (Dynamic Programming). The cost function is defined as the sum of a loss cost by failing and an action cost after the detection. And we present propose the optimal solution and the correctness by numerical calculation. Also we clarify the effectiveness by a numerical experimentation.

Keywords: time series, structural change, optimal stopping problem

I. INTRODUCTION

Change point detection (CPD) problem in time series is to find that a structure of generating data has changed at some time point by some cause. We consider that the problem is very important and that it can be applied to a wide range of application fields [1]. The processing method for the CPD problem is roughly divided into two types: one is batch processing that checks all generated data in the past and another is sequential processing that checks if the structure has changed or not at every new data generation.

As the former representative method, Chow test is well known and is often used in econometrics [2]. It does a statistical test by setting the hypothesis that the change has occurred at time t. As the latter representative method, there are Bayes' method [2],[3] CUSUM, etc., [4] based on sequential probability ratio test.

In practical situation, we have to consider not only that a loss cost is involved with prediction error, but also that an action to be taken after the change detection will need a cost. Conversely, the CPD is necessary in order to judge when to take the action. Under such situations, we must consider the trade-off between loss by the degradation and cost for the quality reformation. However, as far as the authors can know, no such conventional CPD method considering the action cost has been proposed, in spite of the fact that such method is very useful at practical level.

In this paper, we propose a new and practical method based on an evaluation function of loss cost. And we formulate the CPD problem as an optimal stopping problem using the concept of DP (Dynamic Programming) and give the optimum solution by numerical calculation the formulation. We consider that our method is effective in the sense as follows.

1) Differently from the Chow test, it does not need to set the change point in a priori.

2) Unlike the Bayes' method, it does not need to give the generation distribution of time series data.

3) It minimizes the evaluation function that sums up the loss involved with prediction error and action cost to be taken after the change detection.

Also in this paper, we show the effectiveness of our method by numerical experimentation

II. FORMULATION AS OPTIMAL STOPPING PROBLEM

1. Evaluation function

We formulate the CPD problem as an optimal stopping one based on DP with an evaluation function that sums up the cost involved by prediction error and action cost to be taken after the change detection.

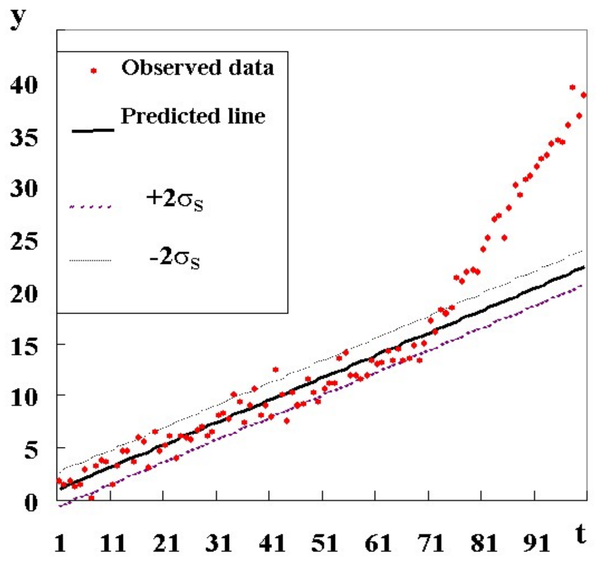


Fig.1. Example of time series data where the true change point tc* =70.

For example, a prediction expression is given in the following equation as a function of time t, where y_t, β_1, β_0, ε mean the function value, two constant coefficients, and error term, respectively.

$$y_t = \beta_1 \cdot t + \beta_0 + \varepsilon \quad (1)$$

The error term ε is given as a random variable of the normal distribution of variance σ^2 and average of 0, i.e., $\varepsilon \sim N(0, \sigma^2)$. A time series data based on the Equation (1) is shown in Fig.1, that is generated by making normal random numbers of average 0 and variance 1 for ε, and by setting β_1=0.2, β_0=1 for the time t=1,2,…,70, and β_1=0.8, β_0=-41 for the time after t=71. The tolerant error interval or tolerance zone between two broken lines as shown in Fig. 1 is decided using the first time series data from t=1 to t=20.

In Fig.1, we think of two situations: one is the situation that the observed data goes out from the tolerance zone, and another the situation that the observed data goes in the zone. We call the former situation "failing" (or "not hitting") and the latter "hitting". We assume that the structure changes when the failing occurs for continuing *N* times.

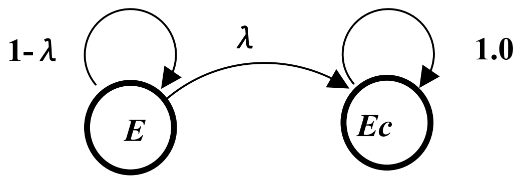
The evaluation function is given by (2) as the sum of two kinds of cost: the damage caused by the failing (i.e., failing loss) and action cost to be taken after the change detection.

$$\text{Total\_cost} = \text{cost}(A) + \text{cost}(n) \quad (2)$$

where cost(n) is the sum of the loss by continuing n times failing before the structural change detection, and cost(A) denotes the cost involved by the action after the change detection. Then we have to find the number of N that minimizes the expectation value of Total_cost, assuming that the structural change occurs randomly.

2. Structural change model

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



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

Fig.2. Structural change model.

Moreover, we introduce a more detailed model. Let *R* be the probability of the failing when the structure is unchanged. Let *R*c be the probability of the failing when the structure change occurred. We consider that *R*c is greater than *R*, i.e., *R*c>*R*. The detailed model for the State *E*c and *E* are illustrated as similar probabilistic finite state automatons in Fig.3 and Fig.4, respectively.

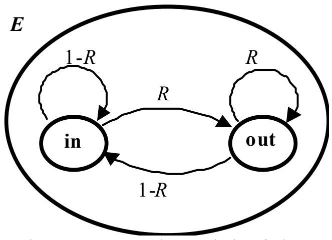


Fig.3. Internal model of the State *E*.

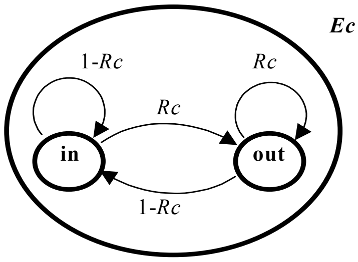


Fig.4. Internal model of the State *E*c.

3. Definition

Let the cost(n) be $a \cdot n$ as a linear function for n, where a is the loss caused by the failing in one time. And for simplicity, let C and A denote the Total_cost and cost (A), respectively. Then, the evaluation function in (2) is denoted as the following equation (3).

$$C = A + a \cdot n \tag{3}$$

We recursively define a function $EC(n,N)$ to obtain the optimum number of times n that minimizes the expectation value of the evaluation function of Equation (3), using the concept of DP (Dynamic Programming). Let N be the optimum number. Let the function $EC(n,N)$ be the expectation value of the evaluation function at the time when the failing has occurred in continuing n times, where n is less than or equal to N, i.e., $0 \le n \le N$.

Thus the function is recursively defined as follows.

$$\text{(if } n = N\text{)} \quad EC(n,N) = A + a \cdot N \tag{4}$$

$$\text{(if } n < N\text{)} \quad EC(n,N) = P(\bar{S}_{n+1} \mid S_n) \cdot a \cdot n + (1 - P(\bar{S}_{n+1} \mid S_n)) EC(n+1,N) \tag{5}$$

where S_n means the state of failing in continuing n times, $\bar{S}_{n+1}$ the state of hitting at the (n+1) th observed data, and $P(\bar{S}_{n+1} \mid S_n)$ means the conditional probability that the state $\bar{S}_{n+1}$ occurs after the state S_n.

The first term in the right-hand side (RHS) of Equation (4) indicates the expectation value of the evaluation function at the time when hitting happens at the (n+1)th data after the continuing n times failing. The second term in the RHS of Equation (5) indicates the expectation value of the evaluation function for the time when failing happens at the (n+1)th data after continuing n times failing.

Then, from the definition of the function $EC(n,N)$, the goal is to find the N that minimizes $EC(0,N)$, because the N is the same as n that minimizes the expectation value of the evaluation function of (4).

4. Minimization of the evaluation function

The analytical solution N that minimizes $EC(0,N)$ can be deduced. The strict proof needs many pages, then we show numerical solution.

The function $EC(0,N)$ is defined by recursive expressions (4) and (5), then $EC(0,N)$ can be computed by recursively. In the process of this computation, $P(\bar{S}_{n+1} \mid S_n)$ can be calculated as follows.

Let E_{cn} be the event that the structural change occurs once during the period of observation in continuing n times. Let $P(E_{cn} \mid S_n)$ be the conditional probability that the E_{cn} happens under the condition that failing has already occurred for continuing n times. Based on the model in Fig.3 and Fig.4,

$$P(\bar{S}_{n+1} \mid S_n) = (1-R)(1 - P(E_{cn} \mid S_n)) + (1 - R_c) P(E_{cn} \mid S_n) \tag{6}$$

Let E be the event that there is no structural change. According to the Bayes' theorem $P(E_{cn} \mid S_n)$ can be represented as (7).

$$P(E_{cn} \mid S_n) = \frac{P(S_n \mid E_{cn}) P(E_{cn})}{P(S_n \mid E_{cn}) P(E_{cn}) + P(S_n \mid E) P(E)} = \frac{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i}}{\sum_{i=0}^{n-1} (1-\lambda)^i R^i \lambda R_c^{n-i} + (1-\lambda)^n R^n} \tag{7}$$

The procedure to calculate $EC(0,N)$ is shown in Fig.5.

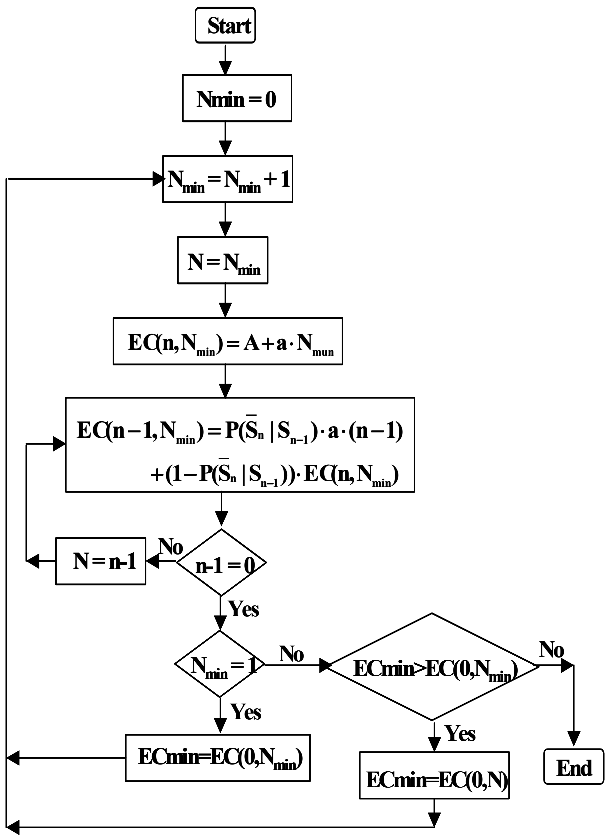


Fig.5. The procedure to calculate N that minimizes $EC(0,N)$ according to the recursive definition.

III. EXPERIMENTATION

In this section, we make a comparison between the proposed method and Chow Test using the time series data shown in Fig.1.

1. Outline of experimentation

Step1: Generate the time series data (Fig.1) based on the Equation (1), by making normal random numbers of average 0 and variance $\sigma^2 = 1$ for ε, and by setting β_1=0.2, β_0=1 for the time t=1,2,…,70, and β_1=b, β_0=(0.2-b)71+1, for the time after t=71.

Step2: Make prediction expression, using a sequence of data at the time t=1…40, from the above generated time series.

Step3: Decide the tolerant zone.

Step4: Based on the proposed method, measure the number of times when the observed data goes out from the tolerance zone for observation data after the time at k+1, and detect the structural change point.

Step5: Perform the above things repeatedly by M times, and calculate the average of the structural change point.

2. Experimental condition

(i)Tolerant zone: $\pm 2\sigma$ of the distribution on error ε.

(ii)The trend β_1 (=b) for the time after t=71:b=0.4.

(iii)Parameter value of the evaluation function:
λ=0.01, and A/a=10,15.

(iv)Repetition times: M=100.

(v)Significance level for testing the hypothesis:
α=0.05 (in case of Chow Test)

3. Results

Results are illustrated in Fig.6 where horizontal axis shows observing time t (observation is started after t=40). The vertical axis in left hand shows an averaged time point of detected change point, where the average is taken from 100 times experimentation. The proposed method detects the change point based on the beforehand calculated value N that minimizes $EC(0,N)$.

Although the detection of the change point depends on the value of A/a, we expect that the change point will be detected around the time at t=70, because the structure of the time series is changed at t=70. We have verified that the results by the proposed method meet our intuition very well.

However, Fig.6 shows that Chow Test decides the change point almost every time when data is observed, and detecting change point varies every time when new data is observed. This means Chow Test cannot correctly detect change point around the time of true change point.

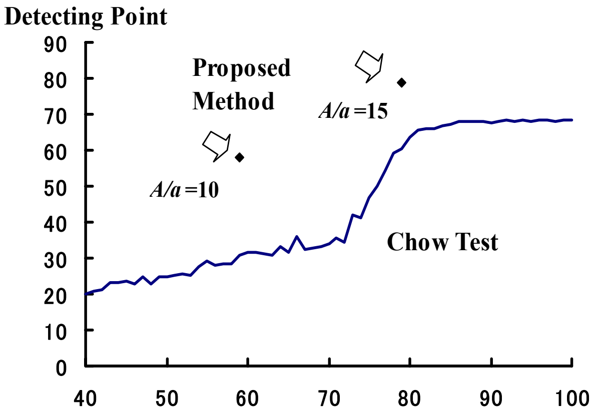


Fig. 6. Change point detection by the proposed method and Chow Test.

IV. CONCLUSION

We have proposed a sequential processing method for structural change point detection of time series data, and have presented a formulation as an optimal stopping problem using the concept of DP. We have defined an evaluation function with an action cost, recursively. In addition, we have revealed the optimum solution by numerical computation and also have shown some experimentation results, where the results meet our intuition and well detect the structural change point in artificially generated time series data. We also have shown that the method is more effective than Chow Test.

We consider that method is effective, especially in the sense that it can quickly detect the change point without a priori knowledge of probabilistic distribution and that it can be applied to arbitrary prediction model, because it is a meta-level one.

REFERENCES

[1] R.Jana and S.Dey (2000), Change detection in Teletraffic Models, IEEE Trans. Signal Processing, vol.48, No.3, pp.846-853.
[2] Chow,G.C. (1960), Tests of Equality Between Sets of Coefficients in Two Linear Regressions, Econometrica, Vol.28, No.3, pp.591-605.
[3] S.MacDougall, A.K. Nandi and R.Chapman (1998), Multisolution and hybrid Bayesian algorithms for automatic detection of change points, Proc. of IEEE Visual Image Signal Processing, vol.145, No.4, pp.280-286.
[4]E.S.Page (1954), Continuous inspection schemes, Biometrika, Vol.41, pp.100-115.