時系列データの記号化による プロセス状態変化検出方式[†]

西 谷 卓 史*

A Symbolizing Method of Time-Series Data for Detecting Signs of a Process Change

Takushi NISHIYA*

This paper presents a symbolization method for detecting process state change from the time series data. The proposed method comprises three processing steps as follows: (a) Peculiarity extracting of process data: the time series process data is expanded into polynomials by a peculiarity extraction filter, and peculiar points are extracted as peak points of the 2nd order polynomial coefficient. (b) Approximating by a series of vectors: the time series data is approximated by a series of vectors (line segments) which connect peculiar points in turn. (c) Comparing with dictionary: the vector series is compared with event patterns which are contained in the dictionary as the vector serieses.

The proposed method has good features as follows: (a) Calculation efficiency: the time series data is processed as a series of vectors, therefore the calculation time in the comparing step is much reduced. (b) Simplicity: the similarity measure is defined as the correlation factor between vector serieses, therefore the patterns which have similar shape but different in size are detected using an event pattern at a time.

The proposed method has been applied to an operation support system for plant and process control. The results prove that the method is practical and able to reduce the operator's work load.

Key Words: symbolization, process data, polynomial expansion, plant operation support

1. はじめに

従来、プラント運転管理システムでのプロセス状態の監視は、データの上下限值や変化率、目標値との偏差などのチェックによって行われていた。この方式は、異常反応、あるいはセンサの故障などデータが大きく変化する異常には有効であり、広く適用されてきた。しかし、実際のプロセスにおいては、センサの不安定化や異常現象による微少な変動など、従来の方式では検出困難な事象変化も発生する。

このような異常状態の見逃しを防ぐために、運転員は

プラントデータの時系列的な変化をトレンドグラフで監視し、異常の兆候を早期に発見できるように努めている。しかし、運転監視の集中化によって、運転員が監視しなければならないデータ数は膨大なものとなり、運転員による人的監視には限度があるという問題があった。重大な事態に至る以前にこのような微少な変化を検出し、運転員に警報を発することができれば、運転員の負荷の軽減ができるばかりでなく、プロセスの安定化にも資することができる。

本論文では、プロセスの時系列データの変化状態を把握し、記号的な言葉に変換する方式を提案する。従来、時系列データのパターン認識の方法としては、空間尺度法に基づくもの⁹⁾や、ニューラルネットを用いるもの¹⁰⁾がある。また、近年ではウェーブレットによる時系列データの解析手法^{8),9)}が脚光をあびている。空間尺度法に基づくものは、パターン特徴を用いたマッチングの手法であ

[†] 第 11 回知識・知能システムシンポジウムで発表 (1990・3)

* (株)日立製作所システム開発研究所 川崎市麻生区
王禅寺 1099

* Systems Development Lab., Hitachi Ltd., Kawasaki
(Received August 25, 1992)
(Revised February 2, 1993)

り、位置のずれや大きさの変化に対しても柔軟に対応できるという特長があり、局所的なパターンに対しては有効であるが、複雑なパターンのマッチングにはアルゴリズム的な処理が必要であるという欠点があった。ニューラルネットを用いた方法は、学習的にパターンの認識ができるという長所はあるものの、認識するパターンの数の増減に対して、再学習が必要である、学習に時間がかかる、といった欠点がある。また、ウェーブレットによるものは、基本的には局所的なフーリエ変換であり、変化の形状まで把握するのは困難であるという問題がある。

本論文では、パターンの特徴抽出と、認識するパターンの記憶部分を分離し、認識するパターンの追加変更に対して、再学習が不要で、柔軟性の高い時系列データの記号化方式を提案する。基本的な考え方は、時系列データを多項式に展開して特徴点を抽出する点にある。特徴点で分割された折れ線ベクトルの系列で表現し、折れ線ベクトルの系列と、あらかじめ折れ線ベクトルの系列として記憶している事象パターンとの間で DP (Dynamic Programming)^{6),7)}を用いてベクトル的な相関係数を計算することにより事象の認識(記号化)を行うことができる。

提案手法には以下のような特長がある。

- (1) 簡単なフィルタリングによって、時系列データの平滑化と多項式展開を同時に実行できる。
- (2) 多項式に展開する特徴抽出フィルタは正規化、直交化を行っているために、展開係数から2次関数で近似した値を特徴点として抽出できる。

- (3) 時系列データを、特徴点を結ぶベクトルの系列として近似して扱うために、事象変化抽出のための辞書との照合処理の演算量が大幅に低減できる。

以下、2章は時系列データの記号化方式の説明であり、2.1節では時系列データの全体の流れ、2.2節では多項式展開の方法、2.3節では特徴点の抽出方法、2.4節では事象パターンとの照合方法、について述べる。3章では実際のプロセスデータによる検証結果により提案手法の有効性を示す。

2. 時系列データの記号化

2.1 記号化処理の流れ

時系列データ記号化方式の全体の流れを Fig. 1 に示す。まず、時刻の初期値を $t=1$ として処理を開始する。また、処理に必要な過去の点数のデータも設定する。つぎに、多項式展開フィルタを用いた積和演算により特徴点の抽出を行う。ここで、2次の展開係数が極値となる点(曲率が極値となる点)を特徴点と呼ぶ。特徴点が抽出されるごとに一つ前に検出された特徴点と新しい特徴点を接続して線分(ベクトル)を構成してゆく。なお、最新のサンプリング時刻は常に特徴点とする。

各時刻において、構成された線分の系列(ベクトル系列)と辞書ファイルにベクトル系列として記憶されている各事象のベクトル系列とを照合し、類似度(Similarity)と尺度(Scale)(事象パターンとの大きさの比率)を計算する。

検出の対象となっている全事象との照合が終了すると、類似度と尺度が一定の値以上の事象があれば、その

Fig. 1 Flow chart of the process state change detection by a symbolization method

事象名が現在のプロセス状態変化を表現しているものとしてアラーム出力される。

以上の処理を各サンプリング時刻ごとに繰り返しながらプロセス状態の監視が行われる。以下、多項式展開、ベクトル系列の生成、および辞書(事象の集合)との照合について詳しく説明する。

2.2 時系列データの多項式展開

時系列データの展開方法として、

$$\int_a^b \{H_m(x)H_n(x)\}E(x)dx=0 \quad (m \neq n) \quad (1)$$

を満足する m 次の多項式 $H_m(x)$ ($m=0, 1, 2, \dots$) を用いる。ここに、 a および b は定義区間を示す定数であり、 $E(x)$ は

$$E(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (2)$$

あるいは

$$E(x) = \exp\left(-\frac{x}{\sigma}\right) \quad (3)$$

などの指数関数とする。このとき、

$$\int_a^b \{H_m(x)\}^2 E(x) dx = A_m \quad (4)$$

とおくと、

$$\phi_m(x) = \frac{1}{\sqrt{A_m}} H_m(x) \{E(x)\}^{1/2} \quad (5)$$

は、正規直交関数系を構成する。

正規直交関数系 $\phi_m(x)$ は、(1)式を満足する多項式 $H_m(x)$ を指数関数 $\{E(x)\}^{1/2}$ によって重み付けたものと考えることができる。時系列データのある時刻を中心として多項式で近似する、すなわち多項式 $H_m(x)$ との相関係数を求めるために、時系列データ $f(t)$ も、時刻 t を中心として、

$$h(t+x) = f(t+x) \{E(x)\}^{1/2} \quad (6)$$

のように、指数関数 $\{E(x)\}^{1/2}$ で重み付けして扱うことにする。

(6)式に対し、(5)式の直交関数系を用いた展開を行うと、

$$\begin{aligned} h(t+x) &= \sum_{m=0}^{\infty} a_m(t) \phi_m(x) \\ &= \sum_{m=0}^{\infty} a_m(t) \frac{H_m(x)}{\sqrt{A_m}} \{E(x)\}^{1/2} \end{aligned} \quad (7)$$

となる。ここに、展開係数 $a_m(t)$ ($m=0, 1, \dots, \infty$) は、

$$a_m(t) = \int_a^b h(t+x) \phi_m(x) dx \quad (8)$$

により求められる。(6)式および(7)式より、時系列データ $f(t)$ は時刻 t の近傍で、(9)式のように

$$f(t+x) = \sum_{m=0}^{\infty} a_m(t) \frac{H_m(x)}{\sqrt{A_m}} \quad (9)$$

多項式に展開されたことになる。

このときの展開係数を求めるフィルタ(以下では特徴

抽出フィルタと呼ぶ) $W_m(x)$ は、(8)式に(5)式および(6)式を代入することにより、

$$a_m(t) = \int_a^b f(t+x) \frac{H_m(x)}{\sqrt{A_m}} E(x) dx \quad (10)$$

となることから、

$$W_m(x) = \frac{H_m(x)}{\sqrt{A_m}} E(x) \quad (11)$$

と表わされる。

以上、連続関数として特徴抽出フィルタ $W_m(x)$ を導いたが、サンプリングデータとの積和演算を行うために、以下では離散的な関数として再定義する。ここで、変数 x は定義区間 (a, b) 内の整数値とし、以下の式における総和演算 $\sum$ は定義区間内の整数値に関する演算を示す。特徴抽出フィルタの具体的な値は、以下のようにして正規化と直交化を行って求められる。

(1) 0 次の項: 0 次の多項式 $H_0(x)$ を

$$H_0(x) = 1.0 \quad (12)$$

とする。正規化定数 A_0 は、(4)式から、

$$A_0 = \sum_{x=a}^b E(x) \quad (13)$$

と求められる。

(2) 1 次の項: 1 次の多項式を

$$H_1(x) = x + c_{10} \quad (14)$$

とおくと、0 次の項との直交条件

$$\sum_{x=a}^b (x + c_{10}) \cdot E(x) = 0.0 \quad (15)$$

から、

$$c_{10} = -\frac{\sum_{x=a}^b x \cdot E(x)}{\sum_{x=a}^b E(x)} \quad (16)$$

と求められる。また、正規化の定数 A_1 は

$$A_1 = \sum_{x=a}^b \{H_1(x)\}^2 \cdot E(x) \quad (17)$$

より求められる。

(3) 2 次の項: 2 次の多項式を

$$H_2(x) = x^2 + c_{21}x + c_{20} \quad (18)$$

とおくと、1 次の項との直交条件

$$\sum_{x=a}^b H_1(x) \cdot H_2(x) \cdot E(x) = 0.0 \quad (19)$$

が成立しなければならない。ここで、(15)式を考慮すると、

$$\sum_{x=a}^b H_1(x) \cdot (x^2 + c_{21}x) \cdot E(x) = 0.0 \quad (20)$$

が成立しなければならないことから、

$$c_{21} = -\frac{\sum_{x=a}^b x^2 \cdot H_1(x) \cdot E(x)}{\sum_{x=a}^b x \cdot H_1(x) \cdot E(x)} \quad (21)$$

と求められる。さらに、0 次の項との直交条件から

$$\sum_{x=a}^b (x^2 + c_{21}x + c_{20}) \cdot E(x) = 0.0 \quad (22)$$

が成立するために,

$$c_{20} = - \frac{\sum_{x=a}^b (x^2 + c_{21}x) \cdot E(x)}{\sum_{x=a}^b E(x)} \quad (23)$$

と求められる. また, 正規化の定数 A_2 は

$$A_2 = \sum_{x=a}^b \{H_2(x)\}^2 \cdot E(x) \quad (24)$$

より求められる.

以下, 同様の計算を繰り返すことによって, さらに高次の多項式も順番に求めることができる. なお, 定義区間が $b = -a$ であり, 指数関数として (2) 式を用いた場合, 多項式 $H_m(x)$ は Hermite の多項式¹⁾となる. また, 定義区間が $b = 0$ で指数関数として (3) 式を用いた場合, 多項式 $H_m(x)$ は Laguerre の多項式¹⁾となる. 本論文では, リアルタイムで時系列データを識別するために, 特徴抽出フィルタの定義区間は $b = 0$ とする.

具体的に, 定義区間を $(a, b) = (-15, 0)$ として特徴抽出フィルタの例を求めた結果を Fig. 2 に示す. Fig. 2 の (a), (b) は, それぞれ (2) 式を指数関数としたときの $\sigma = 2.0$ と $\sigma = 4.0$ の場合を示している. また, (c), (d) は, それぞれ (3) 式を指数関数としたときの $\sigma = 2.0$ と $\sigma = 4.0$ の場合を示している. Fig. 2 からわかるように, 指数関数が (3) 式の場合は, 展開係数への影響が x の大きい部分までおよぶために, ある時刻近傍の変化状況の近似に, 時間的に隔たった過去の影響が大きく出てしまう. したがって, リアルタイムで時系列データの識別を行うには, (2) 式の指数関数を用いるほうが良いと考えられる.

一般的に時系列データは局所的には 2 次式で近似できるため, 特徴抽出フィルタは 2 次までで十分である. ただし, 時系列データの変化が 2 次の多項式で近似できるように, 特徴抽出フィルタの σ を決める必要がある. 具体的には, ある特徴点近傍の近似に, 隣接する特徴点近傍の変化による影響が生じないように, 「 $\sigma =$ 抽出したい特徴点間の最少サンプリング点数」に設定すればよい. ここで, 時系列データからどの点を特徴点として抽出するか(どの程度の変化までをデータの変化として抽出するか)は, データの種類や用途によって, ユーザが決定する必要がある.

特徴抽出フィルタ (11) 式による多項式展開の結果を用いて, 時刻 t における時系列データの変化を多項式で近似した推定値 $f_e(t)$ は, (9) 式に $x = 0$ を代入して,

$$f_e(t) = \sum_{m=0}^{\infty} a_m(t) \frac{H_m(0)}{\sqrt{A_m}}$$

$$(a) \quad E(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right), \sigma = 2.0$$

$$(b) \quad E(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right), \sigma = 4.0$$

$$(c) \quad E(x) = \exp\left(-\frac{x}{\sigma}\right), \sigma = 2.0$$

$$(d) \quad E(x) = \exp\left(-\frac{x}{\sigma}\right), \sigma = 4.0$$

Fig. 2 Examples of the filter for inflection extracting

$$= \sum_{m=0}^{\infty} a_m(t) W_m(0) \quad (25)$$

と求められる。この値は、2.3 節において特徴点を求める計算で使用される。

多項式展開による特徴抽出フィルタを用いて時系列データを 2 次までの多項式に展開した結果と、それによる推定値の変化を Fig. 3 に示す。2 次までの多項式で推定値は十分にもとの変化形状を保っていることを示している。

2.3 特徴点の抽出とベクトル系列の生成

特徴点の抽出とベクトル系列生成処理の基本的な考え方は、

- (a) 多項式展開による 2 次の展開係数が正または負の値をもつ区間の重心を求めることによって、特徴点(時系列データの屈曲点)の時刻を求める。
 - (b) 特徴点を結ぶベクトルの系列として時系列データを近似する。
- からなる。

(a) $\sigma=2.0$

(b) $\sigma=4.0$

Fig. 3 An example of extracting inflection and recovering

(1) 特徴点の抽出

特徴点の抽出の基本的な考え方を Fig. 4 に示す。2 次の展開係数が正または負の値をもつ区間(たとえば、Fig. 4 の A の部分)の重心位置 g_j を (26) 式により求める。

$$g_j = \frac{\sum_{t=t_s}^{t_e} (t \cdot a_2(t))}{\sum_{t=t_s}^{t_e} a_2(t)} \quad (j=0, 1, \dots) \quad (26)$$

ここに、 t_s および t_e は 2 次の展開係数が同じ符号をもつ区間の開始時刻と終了時刻を示す。

つぎに、重心位置 g_j から特徴点を (27) 式によって求める。

$$\tau_j = g_j - c_{21}/2 \quad (27)$$

ここで、右辺の第 2 項は、2 次の多項式(放物線)の中心軸と展開中心の時刻との差である。この補正を行う理由は、多項式展開による 2 次の展開係数は、ガウス関数で重み付けされた 2 次の多項式との相関係数と考えられ、2 次の展開係数の変化に比べ、特徴点の位置は $c_{21}/2$ だけ進んでいるためである。定数 $c_{21}/2$ は定義区間が $(a, b) = (-\infty, 0)$ で連続な場合は σ になるが、Fig. 1 の (a) および (b) のように離散的な場合、それぞれ 1.92 および 4.08 となる。

具体的な特徴点の抽出方法を述べる。2 次の展開係数の値に応じて、つぎのように FLAG を設定する。

$$\begin{aligned} a_2(t) > \Delta: \text{FLAG} &= 1 \\ \Delta \geq a_2(t) \geq -\Delta: \text{FLAG} &= 0 \\ a_2(t) < -\Delta: \text{FLAG} &= -1 \end{aligned} \quad (28)$$

この FLAG の前サンプリング時刻の値からの変化に応じて、特徴点の抽出処理はつぎのように場合分けされる(以下の処理は各サンプリング時刻ごとに実行される)。

- (a) FLAG が反転した場合 (FLAG=1 $\rightarrow$ -1 or -1 $\rightarrow$ 1)

- ① $J \rightarrow J+1$ とインクリメント(特徴点の数が増加)
- ② $J-2$ 番目の特徴点を重心位置から求める(確定)
- ③ $J-1$ 番目の特徴点を重心位置から求める(未確定)

Fig. 4 Method of extracting peculiar

- (b) FLAG が1または-1で継続した場合
 (FLAG=1→-1 or -1→1)
 (FLAG=1→-1 or -1→1)
 ① $J-1$ 番目の特徴点を重心位置から求める(未確定)
- (c) FLAG が0で継続した場合(FLAG=0→0)
 ① なにもしない
- (d) FLAG が0となった場合(FLAG=1→0 or -1→0)
 ① $J-1$ 番目の特徴点を重心位置から求める(確定)
- (e) FLAG が0から変化した場合
 (FLAG=1→0 or -1→0)
 ① $J \rightarrow J+1$ とインクリメント(特徴点の数が増加)
 ② $J-1$ 番目の特徴点を重心位置から求める(未確定)

いずれの場合も、最新のサンプリング時刻($t=T$)は常に特徴点の位置とする。Fig. 4は(b)の場合の特徴点計算の状況を示している。同図のAの部分の重心位置 g_{J-2} から確定している特徴点 τ_{J-2} が求まり、同図のBの部分の重心位置 g_{J-1} から未確定の特徴点 τ_{J-1} が求められる。さらに、現在時刻(最新のサンプリング時刻) T は最新の特徴点 τ_j となる。

求められた各特徴点における時系列データの値としては、(25)式で求められる推定値を用いる。すなわち、時系列データの特徴点は

$$(\tau_j, f_e(\tau_j)) \quad (j=J, J-1, J-2, \dots) \quad (29)$$

と求められる。

(2) ベクトル系列の生成

求まった特徴点の列を順に結ぶベクトルの系列を(30)式によって求める。

$$\begin{aligned} a_j &= (\tau_j - \tau_{j-1}, f_e(\tau_j) - f_e(\tau_{j-1})) \\ (j=1, 2, \dots, J) \end{aligned} \quad (30)$$

特徴点抽出結果の例をFig. 5に示す。同図では、 $\Delta = 0.0001$ としている。2次の展開係数 $a_2(t)$ から得られたベクトル系列 a_j は、時系列データ $f(t)$ の特徴を非常に良く近似できていることがわかる。特徴点の頻度をみると、Fig. 5(a) ($\sigma=2.0$)の場合は時系列データを近似するために必要な以上の特徴点が抽出されている。一方、Fig. 5(b) ($\sigma=4.0$)の場合は、適当な個数の特徴点が抽出されている。この例では、抽出したい変化パターンはtime=60以降の変化であり、抽出したい特徴点間の最少サンプリング点数は4～5である。この図の結果は、2.2節で述べた σ の選択方法が妥当であることを示している。

2.4 辞書との照合

時系列データから抽出する事象の変化パターンも、折れ線ベクトルの系列で表現し、そのベクトル系列を記憶しておく。以下では、これを事象ベクトルと呼び事象ベクトルの系列と記号の組合せの集合を辞書と呼ぶことにする。本節では、辞書の構成を説明した後、事象ベクトルと計測ベクトルの類似度と尺度の定義を行い、それを用いた照合の処理について説明する。

(1) 辞書の構成

時系列データから検出したい記号(事象名)は、事象名辞書として記憶されている。事象名辞書は、事象名に対応した変化パターンをベクトル系列として記憶しており、その例をTable 1に示す。表中、()内の事象名は点線で示した反対事象を示している。事象としては、これら

(a) $\sigma=2.0$

(b) $\sigma=4.0$

Fig. 5 Result of extracting peculiar and constructing vector series

Table 1 Example of dictionary

Fixed →	Step up (Step down) ↗↘	Increasing (Decreasing) ↗↘
Rise up (Fall down) ↗↘	Rise to fixed (Down to fixed) ↗→	Rise and down (Down and rise) ↗↘

以外にも、任意のベクトル系列を登録できる。

(2) 類似度と尺度

類似度および尺度の求め方について述べる。事象名辞書に記憶されている一つの記号のベクトル系列を $\mathbf{s}_i$ ($i=1, 2, \dots, I$)、時系列データに含まれる対応候補のベクトル系列を $\mathbf{a}_j$ ($j=1, 2, \dots, J$) とする。

まず、事象名辞書の各ベクトルと計測データの各ベクトルが 1 対 1 に対応すると仮定する。この場合、事象名辞書のベクトル系列を K 倍したとき、両ベクトル系列の間の 2 乗誤差 E は、

$$E = \sum_{i=1}^I |K \cdot \mathbf{s}_i - \mathbf{a}_i|^2 \quad (31)$$

と表わされる。まず、2 乗誤差 E を最小とする K (以後、尺度と呼ぶ) を求める。(31) 式を展開すると、

$$E = K^2 \sum_{i=1}^I |\mathbf{s}_i|^2 - 2K \sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i) + \sum_{i=1}^I |\mathbf{a}_i|^2 \quad (32)$$

となる。ここに、 $(\cdot)$ はベクトルの内積を示す。(32) 式は K^2 の係数が正の 2 次式であるから、2 乗誤差 E の最小値を与える尺度 K は、

$$\frac{\partial E}{\partial K} = 2 \cdot K \cdot \sum_{i=1}^I |\mathbf{s}_i|^2 - 2 \sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i) \quad (33)$$

より、

$$K = \frac{\sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i)}{\sum_{i=1}^I |\mathbf{s}_i|^2} \quad (34)$$

と与えられる。そのときの 2 乗誤差 E は、(34) 式を (32) 式に代入して、

$$E = \sum_{i=1}^I |\mathbf{a}_i|^2 - \frac{\left\{ \sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i) \right\}^2}{\sum_{i=1}^I |\mathbf{s}_i|^2} \quad (35)$$

で与えられる。(35) 式における 2 乗誤差 E を時系列データのベクトル系列に関して正規化すると、

$$E = 1 - \frac{\left\{ \sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i) \right\}^2}{\sum_{i=1}^I |\mathbf{s}_i|^2 \sum_{i=1}^I |\mathbf{a}_i|^2} \quad (36)$$

となる。したがって、尺度 K により正規化された 2 乗誤差 E による対応度合の評価指標を、(36) 式の第 2 項の平方根を用いた類似度 S で定義する。

$$S = \frac{\sum_{i=1}^I (\mathbf{s}_i, \mathbf{a}_i)}{\sqrt{\left(\sum_{i=1}^I |\mathbf{s}_i|^2 \right) \left(\sum_{i=1}^I |\mathbf{a}_i|^2 \right)}} \quad (37)$$

(36) 式では、(34) 式による尺度 K が正の場合でも負の場合でも 2 乗誤差 E は同じとなる。しかし、時系列データに記号(事象名)を与える場合、反転して一致する場合は類似度は小さくならなければならない。(37) 式は、(36) 式の右辺第 2 項の平方根を考えることにより、適正な類似度を与えられるものとなっている。(37) 式で表わされる

類似度 S は $-1.0 \leq S \leq 1.0$ を満足し、通常の相関係数とよく似た式の形をしており、ベクトル系列間の相関係数と考えることができる。

以上、時系列データのベクトル系列と辞書のベクトル系列が 1 対 1 に対応すると仮定して類似度 S および尺度 K の式を求めた。一般には、Fig. 6 に示すように辞書のベクトル系列と時系列データのベクトル系列との間で一つのベクトルと複数のベクトルとが対応する場合は生ずる。Fig. 6 のグループ I のように計測データの一つのベクトルと辞書の複数のベクトルが対応する場合は、(37) 式および (34) 式に示した類似度 S および尺度 K の分子の $\mathbf{s}_i$ および分母の $|\mathbf{s}_i|$ を

$$\mathbf{s}_i \rightarrow \sum_{k=1}^{K_i} \mathbf{s}_{ik} \quad (38)$$

$$|\mathbf{s}_i| \rightarrow \sum_{k=1}^{K_i} |\mathbf{s}_{ik}| \quad (39)$$

と置き換えて計算する。同様に、Fig. 6 のグループ II のように辞書の一つのベクトルと計測データの複数のベクトルが対応する場合は、(37) 式および (34) 式に示した類似度 S および尺度 K の分子の $\mathbf{a}_i$ および分母の $|\mathbf{a}_i|$ を、

$$\mathbf{a}_i \rightarrow \sum_{l=1}^{L_i} \mathbf{a}_{il} \quad (40)$$

$$|\mathbf{a}_i| \rightarrow \sum_{l=1}^{L_i} |\mathbf{a}_{il}| \quad (41)$$

と置き換えて計算を行う。ここに、 K_i および L_i はそれぞれ i 番目のグループに属する事象名辞書のベクトルの数と時系列データのベクトルの数であり、常に一方は 1 である。

これは、一つのグループ内の複数のベクトルのそれぞれの長さに比例して相手のベクトルを分割した場合に一致する。また、(38), (39) 式または (40), (41) 式に変換することによって、類似度 S は常に $-1.0 \leq S \leq 1.0$ の範囲に保たれる。

(3) 辞書との照合

辞書との照合の流れを Fig. 7 に示す。時系列データに対するベクトル系列と、事象のベクトル系列とを DP (Dynamic Programming) を用いて照合する。辞書の事象に対して、類似度と尺度が

$$\begin{cases} S \geq \underline{S} & (\underline{S}: \text{類似度の下限値}) \\ K \geq \underline{K} & (\underline{K}: \text{尺度の下限値}) \end{cases} \quad (42)$$

Fig. 6 Groups of matched vectors

Fig. 7 Flow chart of DP matching

を満足する事象があれば、その事象のベクトル系列と時系列データのベクトル系列は一致したと判定する。

照合に用いる経路を Fig. 8 に示す。横軸方向は辞書の事象ベクトル s_i ($i=1, 2, \dots, I$)、縦軸方向は時系列データのベクトル系列 a_j ($j=1, 2, \dots, J$) であり、○は両者の対応関数を表すノードである。DP による照合は、時間方向の入れ替わりがないとすると、対応関係は Fig. 8 のように、縦、横および右上がりの斜めのアークで接続された経路から最も類似度の高い経路を計算することになる。

尺度が一定の値の範囲 ($K_{\min} \leq K \leq K_{\max}$) をとりうる場合のみ照合を行うことにすると、 $(1, j)$ のノードのなかで対応関係の開始点となりうるノードの番号 j ($j=j_s \sim j_e$) は、

$$\begin{cases} j_s = \left\{ \min_j j \mid \sum_{i=1}^j |a_j| \leq K_{\max} \cdot \sum_{i=1}^j |s_i| \right\} \\ j_e = \left\{ \max_j j \mid \sum_{i=1}^j |a_j| \geq K_{\min} \cdot \sum_{i=1}^j |s_i| \right\} \end{cases} \quad (43)$$

として求められる。各事象のベクトル系列との照合は j_s 番目のベクトル以降で行えばよいことになり、演算量の増加を抑えながら、適切な照合を行うことができる。

実際の計算順序は種々考えられるが、ここでは実時間処理を考慮して、 j が一定の (i, j) によって第 j 段を構成し、第 j 段内では i が小さいほうから順次計算し、その後で第 $j+1$ 段に移る方法で計算を行う。このとき、ノード (i, j) に至る可能な経路は Fig. 9 に示した 7 通りがある。これは、縦→横あるいは横→縦が継続する対応関係は多重対応となってしまうために除外した結果である。類似度の計算は (37) 式で行うために、漸化式の形で表現することができない。したがって、ノード (i, j) における

Fig. 8 Map for the DP matching

Fig. 9 Route candidates to the node (i, j)

経路の探索は、Fig. 9 に示した各 3 方向からの経路について別々に計算を行う。すなわち、

$$\begin{cases} r_0(i, j) = d(i-1, j) + r_k(i-1, j) \\ r_1(i, j) = d(i-1, j-1) + r_k(i-1, j-1) \\ r_2(i, j) = d(i, j-1) + r_k(i, j-1) \end{cases} \quad (44)$$

ここに、+ は経路を接続することを表わし、 $d(i-m, j-n)$ はノード $(i-m, j-n)$ からノード (i, j) への経路を表わし、 $r_k(i-m, j-n)$ は k 方向 ($k=0$: 横, $k=1$: 斜, $k=2$: 縦) からノード $(i-m, j-n)$ に至るルートを示している。方向 k は、Fig. 9 に示した各方向に対して、その経路に沿って類似度 S を計算し、最大の類似度となる経路の番号とする。

以上の処理をノード (i, j) まで行い、 $r_k(I, J)$ ($k=0, 1, 2$) のうち最大の類似度をもつ経路が対応関係を表わすことになる。

3. 実データによる検証結果

前章で説明した方式による記号化実験結果を Fig. 10 に示す。照合に用いた事象のベクトル系列は Table 1 に示した事象と、「Fixed」以外の事象の変化方向を反転した事象の 11 個を用いた。照合範囲を定める尺度の範囲は $K_{\max}=2.0$, $K_{\min}=1.0$ とし、類似度の下限値 $\underline{S}=0.9$, 尺度の下限値 $\underline{K}=1.0$ 以上の場合に事象が成立したとしている。各事象が成立した時点を示し、事象ベクトル

Fig. 10 Examples of the process state change detection

ルの系列と一致した区間を—で表示している。また、連続したサンプリング時刻で同一の事象が継続して成立した場合は表示を行っていない。いずれの場合も、時系列データの変化に従って、その変化形状を正しく表現する事象名(記号)が得られている。Fig. 10 のような連続プロセスのデータでは、事象「Fixed」が頻繁に発生する。このような場合は、事象「Fixed」は定常状態として、検出対象の事象から除外すればよい。また、検出事象の数が多すぎる場合は、重複して検出される事象も辞書から除外すればよい。たとえば、Fig. 10(a) の場合、「Rise up」「Rise and down」「Decreasing」の順に事象が検出されている。この場合、「Rise and down」は辞書から除外できる。このように、辞書に含まれる事象を自由に登録・削除できる点に本手法の大きな特徴がある。また、検出のための類似度と尺度の限界を事象ごとに定めることも可能である。

Fig. 10(a) において、事象名「Rise up」および「Rise and down」に対する類似度と尺度の変化を Fig. 11 に示す。また、それぞれの事象が成立したサンプリング時刻における事象ベクトルとプロセスデータのベクトル系列との対応状況を Fig. 12 に示す。対応範囲の尺度の下限 $K_{\min}$ を 1.0 としているにもかかわらず、尺度が 1.0 以下の場合が多いのは、類似度が低い場合は(34)式の尺度も小さな値となるためである。類似度が大きい時刻では尺度はほぼ 1.0 以上となっている。また、Fig. 12 に示した

Fig. 11 An example of the similarity and scale

Fig. 12 Examples of the DP matching result

ように、事象とプロセスデータのベクトルデータの間でよい対応関係が得られており、DP マッチングが良好に行われていることを示している。

提案した手法を制御用計算機 HIDIC V90/45 (5.0 MIPS) に実装し、処理速度の検討を行った結果、CPU 負荷 10% の状態で 300 データ/分の処理性能となることを確認した。ここで、辞書に登録した事象は 22 個で平均 5 個のベクトルで構成されており、処理に用いたデータか

ら生成されたベクトル数は平均 10 個であった。この結果、サンプリング間隔が 1 分のデータを監視するには、実用上十分な計算速度であることを実証することができた。

4. おわりに

多項式展開による特徴点の抽出方法を用いた、時系列データの記号化方式について報告した。本方式によれば時系列データを人間が理解できる記号的な表現に変換できる。実データによる検証の結果、実用に十分な結果を得られ、プラント運転員の負荷軽減に大いに役立つと考えている。

本方式は、データベースなどに蓄えられたデータを、変化パターンごとに分類するデータ分類や、蓄えられたデータから一定のパターンで変化したデータを抽出する類似データの検索にも適用できると考えている。今後は、大量データの記号化による分類・検索方式の開発を行う予定である。

最後に、本研究の機会を与えてくださった当所堂免信義所長ほか関係各位、および本研究の遂行にあたって多大なご協力を頂いた同社機電事業部、大みか工場の各位に対してここに感謝の意を表します。

参考文献

- 1) 寺澤寛一：数学概論，岩波書店（1954）
- 2) 西谷，ほか：時系列データの記号化によるプロセス状態識別方式の提案，第 11 回知識・知能システムシンポジウム（1990）
- 3) 小河，ほか：時系列データの記号化によるプロセス状態変化検出システム IMARK の開発，日立評論，73-8（1991）
- 4) Kawaguchi, et al. : Intelligent Plant Operation Support System for Plant and Process Control, Hitachi Review, 41-1（1992）
- 5) Ohta, et al. : Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7-2, 139/154（1985）
- 6) 和田，ほか：周期波形の尺度空間フィルタリング，電子情報通信学会論文誌 D-II，J73-D-II-4, 554/552（1990）
- 7) 追江，ほか：動的計画法を利用した音声の時間正規化に基づく連続単語認識，日本音響学会誌，27-9, 483/490（1971）
- 8) 山口，ほか：ウェーブレットとその応用について，応用数理，1-3, 2/13（1991）
- 9) 山口，ほか：ウェーブレット解析，科学，60-6, 398/405（1990）
- 10) 福見，ほか：ニューラルネットワークによる位置ずれ・回転に不変なパターン認識システム，電気学会論文誌 C，110-3, 148/155（1990）

[著者紹介]

西谷 卓 吏 (正会員)

1973 年，京都大学工学部電気工学科卒業。同年(株)日立製作所に入社。現在，同社システム開発研究所において上下水道などの監視制御システム，プラント運転支援システム，FA システムなどの研究に従事。IEEE，システム制御情報学会，情報処理学会などの会員。