

非線形変換を利用した時系列データからの知識発見

Knowledge Discovery from Time-series Data using Nonlinear Transformations

井手 剛*
Tsuyoshi Idé

井上恵介†
Keisuke Inoue

Abstract: Most of the stream mining techniques presented so far have mainly paid attention to discovering association rules by direct comparisons between time-series data sets. However, their utility is very limited for heterogeneous systems, where time series of various types (discrete, continuous, oscillatory, noisy, etc) acgt dynamically in a strongly correlated manner. In this paper, we introduce a new nonlinear transformation, singular spectrum transformation (SST), to address the problem of knowledge discovery of causal relationships from a set of time series. SST is a transformation that transforms a time series into the probability density function that represents a chance to observe some patricular change. For an automobile data set, we show that SST can discover a hidden and useful relationship between variables.

1 はじめに

不規則な変動を含む時系列データは古くから研究者の興味を引いてきた。フーリエ解析やウェーブレット解析、あるいは自己回帰モデルなど、時系列の理解と特徴づけのために、ありとあらゆる試みが今までなされてきた [14]。その長く深い研究の歴史にもかかわらず、近年、データマイニングの文脈から、時系列からの知識発見が新たな研究領域として再認識されている。

もともとデータマイニングは、多数のアイテム間の相互関係を見出すことを目的として進歩してきた [15]。その問題意識の自然な拡張として、複数時系列の相互関係の記述やそこからの知識発見が、時系列マイニングの主要なテーマとなると考えるのは自然である。そしてそこそが、時系列解析の研究分野に残されたひとつの空白地帯であると我々は考えている。

時系列データ源として、コンピュータシステムや、機械システムを想像しよう。我々はそういったシステムを、動的でヘテロな強相関係として捉える。すなわち、本論文で想定するシステムは、離散・連続・振動的・ノイジー、などさまざまな時系列データを時々刻々生み出す。それ

らは何らかの関連を持って時間変動し、系全体の挙動は、個々の変数の単純な和としては理解できない。

一般に動的な強相関係では、個々の変数対の絡み合いの機構が比較的単純であっても、系全体としての振る舞いは非常に複雑になりえる。そのため、そこからの知識発見は想像以上に難しい。たとえば自動車において、エンジン回転数、ギア位置、燃料流量、スロットル位置、吸気酸素濃度、などのセンサー情報は、自動車の置かれた状況、操作の様態によって、ほとんどの無限のパターンを持つ。そのため、異常を呈している自動車の時系列情報から、「変数 x_1 がこの値のとき x_2 がこれ」といった類の規則を見出すのは一般に難しい。

本論文で我々は、時系列データ群からの隠れた因果関係の発見という問題を考える。対象とする動的強相関係の振る舞いの詳細は、事前には知りえないと仮定する。また、各時系列は突発的かつ急峻な変化を含み、トレンド成分と雑音成分に分離といった伝統的な時系列解析の対象にはなりにくい。この問題設定は、時系列マイニングの伝統的なそれ [3] とはやや異なることに注意する。Das ら [3] のルール発見アルゴリズムの暗黙の前提は、時系列の部分列が比較的少数のパターンにクラスタリングされるということであった。比較的素朴な変化をする少数の時系列群は別にして、一般には彼らの手法の有効性はかなり限定的である。また、恣意的に短い部分列を

*IBM 東京基礎研究所, 242-8502 神奈川県大和市下鶴間 1623-14, e-mail: goodidea@jp.ibm.com.

IBM Research, Tokyo Research Laboratory, 1623-14 Shimotsuruma, Yamato-shi, Kanagawa, Japan, 242-8502.

†同上, e-mail: inouek@jp.ibm.com

切り取り、それをひとつのカテゴリに帰着させることの理論的正当性も最近議論の対象となっている [8]。

伝統的な時系列マイニングと同様、我々の問題に対するひとつの有力なアプローチは、時系列データのクラスタリングを行うことである。たとえば、系の正常な動作時にある自然なクラスタリング結果が存在するとすれば、それからのずれを検出することにより、異常の検知ができる。また、逆に、何の異常が観測されたとすれば、自然なクラスタリング結果と照合することにより、その原因をある程度限定することもできよう。ここで重要なのは、動的強相関系に内在する因果関係を正しくクラスタリング結果に反映しなければならないということである。

時系列クラスタリングは、通常、時系列同士の距離を求めるステップと、その距離を元にクラスタリングを行うステップに2分割できる。後者は階層クラスタリングや k 平均法などのいわば「枯れた」技術が存在しているが、前者は必ずしもそうではない。むしろ、因果関係発見という視点から見たとき、従来の時系列マイニングの研究は少なくとも2つの点で必ずしも満足いくものではない。

ひとつには、従来の研究の多くが、時間軸方向の定常性を暗に仮定していたことがある。これについては、Keogh らの批判に詳しい [8]。彼らの得た結論は、時間並進に必ずしも不変でない時系列に対し時間並進不変性を仮定したクラスタリングアルゴリズムを適用すると、その結果が入力時系列に依存しなくなる（つまり結果が信用できない）、というものである。ヘテロな強相関系の代表として自動車の生成する時系列データ群を考えると、それらは決して定常ではありえず、常に何かの変化に見舞われており、それを無視するような手法は利用すべきではない、ということである。先走ってさらに言えば、その変化こそが、変数の分類のキーとなるはずだ、ということである。

既存手法の問題のもうひとつは、いわば空間方向の一様性を暗に仮定していることである。たとえば10個の時系列データを考える場合、その10個の変数がいずれも似たような変動をする変数であると暗に仮定されていることである。この限界を一步踏み越える試みの代表例が、DTW (dynamic time warping) 法を用いた一連の研究である [2, 9]。しかしながらヘテロな強相関系の解析におけるDTW法ないしその派生物の有効性は極めて限定的である。物理的意味も値域も変動の様態も異なる時系列変数同士は、仮に局所的に時間軸の伸縮を行ったとしても、直接の比較が意味を持つとは期待できないからである。たとえば自動車のデータにおいては、エンジ

ン回転数とギア位置には何らかの因果関係があるだろうが、それらの関係について何か述べる手段は、少なくともデータマイニングのコミュニティには存在しなかった。

我々は本論文で、上記2つの困難を、特異スペクトル変換という新しい非線形変換を導入することで乗り越えようと試みる。特異スペクトル変換は、各時系列を、その変化度スコアの時系列に直す変換である。変化度スコアが高い位置は、その変数に「何かがあった」時刻を示している。それゆえ、その位置は、変数の因果関係を指し示す重要なキーとなりえる。また、物理的意味も値域も変動の様態も異なる変数群にこの変換を施すことで、ヘテロな強相関系の時系列マイニングを、従来の時系列クラスタリングの問題に帰着させることができる。

以下、第2章で我々の提案するノンパラメトリックな変化点検出手法について説明する。第3章で新しい非線形変換技術をまとめる。第4章で簡単な実験結果について報告する。第6章でまとめを行う。

2 変化点検出手法

2.1 従来手法の概観

変化点検出問題とは、観測される時系列データの背後にあるデータ生成機構を想定した際、その生成機構の構造的な変化を検出する問題である。単なる外れ値と異なり、段差、カスプ、周波数変化、などさまざまな出現形態をとりえる。その問題自体は比較的古くから知られており、たとえば、ある一定値の周りにデータが分布するような1変数データにかんしては、CUSUMと呼ばれる手法が簡便に使われる [1]。しかしヘテロな強相関系では、定数に回帰するようなモデルを仮定することはできない。

その他代表的な手法としては、自己回帰モデルに基づいて変化点検出を行う方法が知られている。Yamanishi-Takeuchi [13] は、混合正規分布の期待値最大化アルゴリズムを忘却率つきでオンライン拡張したものと、AR (autoregressive) モデルの Yule-Walker 推定式 [14] やはり忘却率つきでオンライン拡張したものを組み合わせ、変化点検出を行う手法を提案している。混合正規分布は基本的に独立同一試行を前提とし、また、自己回帰モデルはデータの定常性を仮定しているが、それらに過去データの忘却効果を取り入れることにより、ある程度適応的に変化点検出を行うことができる。

しかし人間の不規則的な操作が絡むような変動に対しては、自己回帰モデルはよい出発点ではないし、独立同一試行の連鎖として現象を捉えるのも難しい。加えて、我々が興味を持つようなヘテロなデータセットに対し、

パラメータ選択などにおいて、混合正規分布の推定アルゴリズムは実用上いくつかの困難を抱えている。たとえば、ほとんど定数であるような不活性な変数の混入に対して頑強とは言えない上、混合数に対して結果の収束を保証するのが難しい。要するにその枠組みを使うには、ある程度事前知識を使って入念な下準備をする必要がある。これはパラメトリックなモデルを前提とする場合ある程度避けられないものであるが、我々の用途には好ましいものではない。

Hirano-Tsumoto[6] は核関数法による分布推定により任意の粒度で時系列を表し、その時系列の変曲点で部分系列を定義する手法を提案している。いわば変曲点を変化点と見ているわけである。そして、あるヒューリスティックな尺度により、複数の時系列同士でその部分系列を比較することで、時系列の分類などをより正確に行うことができると報告している。彼らの区分法はその発想において非常に進歩性の高いものであるが、2階微分を变化度尺度として採用できる時系列にはかなり強い制約があり、実際、われわれの考えるヘテロなデータセットには適しているとは言えない。

最近、地球物理学の分野で、特異スペクトル解析法 (singular spectrum analysis) という手法が知られるようになってきた (たとえば [5] 参照)。これは時系列データからノイズを取り除くために、時系列から作られるハンケル行列に特異値分解を施し、特徴抽出するものである。ごく最近、Moskvina-Zhigljavsky[11] は、特異スペクトル解析を变化点検出に応用する手法を提案している。彼らの手法はノンパラメトリックかつ数値的にも安定であり、我々の用途に適している。下記に、特異スペクトル解析法を詳しく説明し、我々の用途に適する变化度指標の定義を行う。

2.2 過去の代表パターンの抽出

時系列 $\mathcal{T} = \{x(1), x(2), \dots, x(t), \dots\}$ を考える。以下 t を時刻を表す指標として用いる。 \mathcal{T} の部分列として、長さ w の $\{x(t-w), \dots, x(t-2), x(t-1)\}$ というものを考え、対応する列ベクトルを

$$s(t-1) = (x(t-w), \dots, x(t-1))^T$$

と書く。上付きの T は転置を表す。この列ベクトルを並べて下記のハンケル行列を構成する (図 1 参照)。

$$H(t) = [s(t-n), \dots, s(t-2), s(t-1)].$$

この $w \times n$ 行列を、Moskvina-Zhigljavsky[11] にならい、時刻 t における履歴行列 (trajectory matrix) と呼ぶ。定義から履歴行列は、 $x(t-1)$ から $x(t-w-n+1)$ ま

での $w+n-1$ 個の要素で定義される。この $w+n-1$ を特に W と表しておく。

さて、行列 $H(t)$ は、時刻 t から過去 W の範囲において、長さ w の制約の下、この時系列の様々な変動パターンを列ベクトルとして並べたものと考えることができる。そこで、この $H(t)$ から、この時系列の代表パターンを抽出することを考えよう。代表パターンを u と表し、 $u^T u = 1$ という制約を課しておく。もっとも簡単な仮定は、これが $s(t_j)$ の線形和

$$u = c \sum_{i=1}^n v_i s(t-i)$$

から成るとするものである (c は規格化条件を満たすための定数)。ここで n 次元ベクトル $v = (v_1, \dots, v_n)^T$ を定義すると、この式は簡明に $u = cH(t)v$ と書ける。代表パターンの抽出にはさまざまな方法がありえるが、ここでは各パターンについての上記の一次結合が、もっとも「長い」ベクトルを与えるように、

$$v(t) \equiv \arg \max_{\tilde{v}} \|H(t)\tilde{v}\|^2 \quad (1)$$

という方程式を考える。ただし、 $v^T v = 1$ という制約を置く。ラグランジュ乗数 λ を導入して、この極値方程式は、

$$\frac{\partial}{\partial \tilde{v}} [\tilde{v}^T H(t)^T H(t) \tilde{v} + \lambda \tilde{v}^T \tilde{v}] = 0$$

と変換される。これより直ちに v が固有値方程式

$$H(t)^T H(t)v = \lambda v$$

の規格化された解であることが分かる。また、 u が、 $H(t)H(t)^T$ の固有ベクトルであること、すなわち

$$H(t)H(t)^T u = \lambda u \quad (2)$$

が成り立つことも容易にわかる。これらの結果を言い換えれば、代表パターン u およびその係数の作るベクトル v は、それぞれ、行列 $H(t)$ の左特異ベクトルと右特異ベクトルになっている [16]。その特異値は $\sqrt{\lambda}$ に等しい。

代表パターンとその特異値を、特異値の大きい順に $\{(\sigma_1, u_1), (\sigma_2, u_2), \dots, (\sigma_l, u_l)\}$ と置く。パラメータ l は考慮するパターンの個数であり、特異値の大きいほど主要なパターンで、特異値の小さいものはノイズとみなすことができる。

上記のように、ハンケル行列の特異値分解を経由して時系列から代表パターンを見出す手法を、特異スペクトル解析と呼ぶ。¹

¹ 「スペクトル」という名前がついているが、古典的なフーリエ解析とは全くの別物であることを強調しておく。

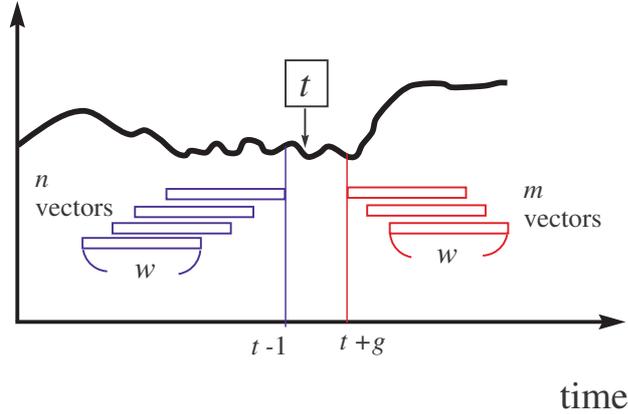


図 1: 特異スペクトル変換で用いる定数の説明。現在時刻の両側にハンケル行列を取る。

2.3 現在の代表パターンの抽出

過去の代表パターンが見出されれば、それと現在のパターンを比較することにより変化度を計算することができるはずである。今、履歴行列の未来側に、再び長さ w の列ベクトル

$$\mathbf{r}(t+g) = (x(t+g), \dots, x(t+g+w-1))^T$$

をとる。これは $\mathbf{s}(t+g+w-1)$ と同じものである。履歴行列と同様に、今考えている時刻 t における「テスト行列」を、 m 本の列ベクトルを用いて

$$G(t) = [\mathbf{r}(t+g), \mathbf{r}(t+g+1), \dots, \mathbf{r}(t+g+m-1)]$$

で定義しておく。先ほどと同様に考えれば、現在の代表パターンは、式 (2) と同様に、

$$G(t)G(t)^T \mathbf{u} = \lambda \mathbf{u} \quad (3)$$

の解として求められる。最大固有値に属する規格化された固有ベクトルを特にテストベクトルと呼び、 $\beta(t)$ と表す。

2.4 変化度の定義

さてわれわれは、過去の変化パターンとして、 $\{\mathbf{u}_i | i = 1, \dots, l\}$ を、また、現在時刻近傍の代表的な変化パターンとして $\beta(t)$ を得た。これらから、異常度を定義することを考えよう。もしも $\beta(t)$ が過去のパターンをなぞったものであれば、 $\{\mathbf{u}_i | i = 1, \dots, l\}$ の張る超平面内にあるだろう。しかしもし何らかの変化が生じていればその超平面からはみ出すことがありえる。そこで、その「はみ出し」を定量的に評価することを考える。

過去の代表パターンを並べた行列 U_l を、

$$U_l = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l]$$

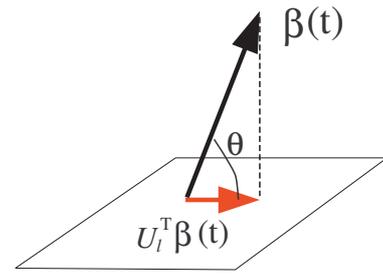


図 2: 特異ベクトルで張られる超平面への射影。角度 θ が大きいほど変化度が大きいことになる。

と書く。すると、テストベクトルの超平面への規格化された射影は、

$$\alpha(t) \equiv \frac{U_l^T \beta(t)}{\|U_l^T \beta(t)\|}$$

と書ける。もしテストベクトルが超平面と直交していればゼロベクトルとする。すると、図 2 に示すように、

$$z(t) \equiv 1 - \alpha(t)^T \beta(t) \quad (4)$$

によって変化度を定義することができる。定義より明らかに、この量は、全く変化がないときにゼロ、変化が最も激しい時(すなわち α が超平面と直交しているとき)に 1 という値をとる。

なお、Moskvina-Zhigljavsky[11] は、角度ではなくノルムの変化について異常度を定義しているため、異常度が一般に無次元量にならない。また、テスト行列からのパターン抽出は考慮されておらず、変化度スコアに対して理論的根拠の不明確な平均操作が行われている。われわれの定式化には、 $l=1$ かつ $m=n$ の時に、時間軸の反転 ($t \rightarrow -t$) に対し対称であるという特徴がある。時系列を時間反転させて得られる変化度スコアは反転前とほぼ一致すべきである、という直感的要請からすれば、より自然な定式化であると思われる。

この変化点検出手法の特徴は、アルゴリズムの中核となる特異値分解という操作が任意の行列に対して定義され、それゆえ、多彩な入力ベクトルに対して汎用に見えることである。実際、工作機械の時系列データで言えば、モードレバー位置のような離散変数と、回転速度のような連続変数に対して共通のアルゴリズム・共通のパラメータで変化度の計算をすることが可能である。ヘテロな動的システムに非常に適した手法であるといえる。

3 特異スペクトル変換

3.1 定義

上記のように、履歴行列とテスト行列におけるパターンを逐次計算していくことにより、各時刻 t において変

化度を計算することができる。今、ある時系列 T から、変化度スコア $z(t)$ の時系列 T_c への変換

$$T \rightarrow T_c(w, g, l)$$

を特異スペクトル変換と定義する。括弧内はこの変換で用いる主なパラメータを表している。この変換は、重ね合わせの原理を満たさない等の意味において、時系列の非線形変換を定義している。変換後の時系列について、以降、積分面積を 1 に規格化するものと約束する。これにより、変換後の時系列は、「変化が生ずる確率密度」として解釈することができる。時系列に起こる変化点の発生は、離散的・連続的・ノイズ・振動的、等の表面的相違とは独立なはずである。すなわち、新しい時系列 $T_c(w, g, l)$ は、元の時系列の表面的な振る舞いの陰に隠れた因果律の痕跡を表すと考えられる。したがって、変換後の時系列につき互いの類似度を議論することは、システムの因果律の状態を論ずる上で有力な手段となりえる。

3.2 パラメータの選択

特異スペクトル変換において探知できる変化点の時間スケールを定める重要なパラメータが w である。変化点の発生間隔が w より小さいと探知は難しいことに注意する必要がある。たとえば平均して 1 分間に 1 度シフトチェンジが行われるような状況であれば w は 1 分に対応する値よりも十分小さく設定されるべきである。

履歴行列の列の数 n についてはしばしば w と等置される。テストベクトルの列の数 m に計算は敏感ではなく、1 から w の範囲で適当に決めてよい。 g については、直感的に、テスト行列の及ぶ範囲の中心がちょうど t あたりになるように、 $-(w+m)/2$ 程度にするのがよいと考えられる。

履歴行列の左特異ベクトルの本数を表すパラメータ l は、通常の主成分分析でそうするように、特異値の寄与率から見積もることができる。すなわち、式 (2) において、直交変換でトレースが不変であることを考えれば、

$$r_l \equiv \frac{\sum_{i=1}^l \sigma_i^2}{\text{trace}[H(t)H(t)^T]}$$

を、 l 番目の特異値までの寄与率と考えることができる。通常、数個程度の l を取ったあとは、寄与率の増分が急減するため、 r_l の l 依存性を見ると容易にノイズと主成分を分離することができる。

4 実験

実験で用いるデータセットを図 3(a) に示す。これはある機械システムを模したテストデータで、0.1 秒間隔

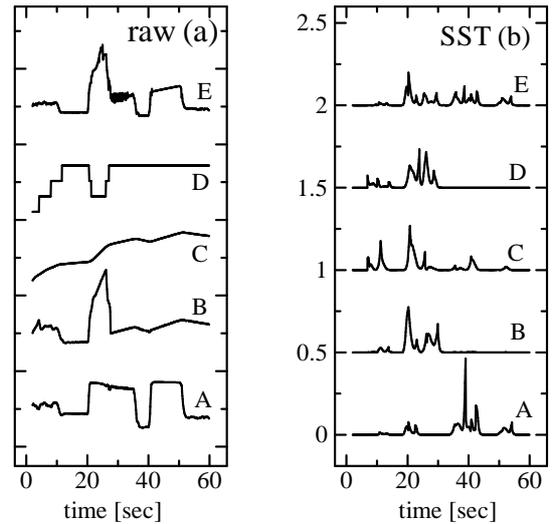


図 3: ある機械システムの時系列データとそれに SST を施した結果。(a) 元の時系列データ。(b) SST の結果。

で約 1 分間のデータである。見やすいように、B から E に関しては適宜上に平行移動してある。一見して分かるとおり、このデータセットには、離散的なもの (D)、なだらかに変動するもの (C)、急峻な変化をするもの (A, B)、ノイズなもの (E) と、様々な様態のデータが混合されている。

この時系列データに関し、SST を施したものを図 3(b) に示す。結果は積分面積を 1 に規格化してあり、見やすいように上下方向に 0.5 ずつずらして表示した。パラメータは $w, m, n, -g$ いずれも 2.5 秒である。採用する過去パターン数は $l=2$ とした。このパラメータの下、計算時間はひとつの時系列につき 1 秒未満であった²。SST を行うにあたり、元データは標準化変換の後 3 を加えたものを使用した。

元データの図と見比べることにより、各ピーク位置が、大きな変化に対応していることが見て取れる。興味深いのは、C のようななだらかな変数と、E のような激しい変動をする変数の変化点が同期しているように見えることである。また、B と D は生の時系列では似ても似つかないが、SST により共通性が浮かび上がっていることがわかる。すなわち、元のデータでは各時系列の挙動はあまりに多様で、それらを直接比較するのは難しいが、SST という非線形変換を施した後は、互いに比較可能になっていることが見て取れる。このようにヘテロな時系列を「同じ土俵に乗せる」ことが、SST の重要なご利益である。

²Java で実装された IBM の数値計算ライブラリを、標準的な PC 上で使用。

5 考察

5.1 時系列同士の距離尺度の定義

図3の結果から、定量的知見を引き出すために、時系列変数同士の距離を計算してみよう。SSTの確率解釈に従えば、変化度の時系列 $z_i(t)$ と $z_j(t)$ の自然な距離尺度は、Hellinger距離

$$d_H(i, j)^2 = \int dt \left[\sqrt{z_i(t)} - \sqrt{z_j(t)} \right]^2$$

である。Hellinger距離は、Bhattacharyya類似度 $a_B(i, j)$ と $d_H(i, j)^2 = 2 - 2a_B(i, j)$ で関係付けられ[10]、各変数をグラフのノードと見て類似度グラフのスペクトルクラスタリング[4]を行う際にも便利である。

距離を計算するにあたり、下記の式によりSSTされた時系列 $z_i(\tau)$ を変換する。

$$Z_i(t) = \int d\tau z_i(\tau) \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{(t-\tau)^2}{2\sigma^2} \right] \quad (5)$$

これは直感的には、各ピークを幅 σ の正規分布で「拡幅」したことに当たる。この操作は、第1に、測定における時間方向の誤差を考慮したという意味を持つ。第2に、観測データに含まれる見かけ上のノイズを濾過して、「変化が生じた確率」としてのSST時系列を純化させたという意味を持つ。すなわち、機械システムの本質的な変化点は、たとえば人間の操作により引き起こされるため、それがノイズな振舞いをするのは不自然である。また、Hirano-Tsumoto[6]流に言えば、 σ をさまざまに選ぶことで、着目する時間粒度を多重スケール化しているとも考えることができる。図4に、図3を $\sigma = 1.5$ [秒]により平滑化した結果を示す。

5.2 多次元尺度構成法による可視化

多次元尺度構成法 (multidimensional scaling, MDS) は、与えられた距離行列から元の座標を復元する手法である[10]。復元する座標の次元はパラメータとして選択できるが、通常2次元が好んで用いられる。ここではいわゆるMDSの古典解(補遺参照)に従って、図4におけるAからEの5つの変数を、2次元平面に散布図プロットした。結果を図5に示す。SSTに関しては上記のHellinger距離を用い、元の時系列に関しては、標準化変換を行ってから(すなわち平均ゼロ・標準偏差1にしてから)、2乗積分面積を1に規格化し、その後 L_2 距離により距離行列を構成した³。

なお、MDSの性質から、復元された座標の重心はゼロで(したがって原点周りに分布し)、また、直交変換

³Hellinger距離を使わなかったのは、(1乗)積分面積が1に規格化されていないこと、時系列が正にも負にもなりえることによる。

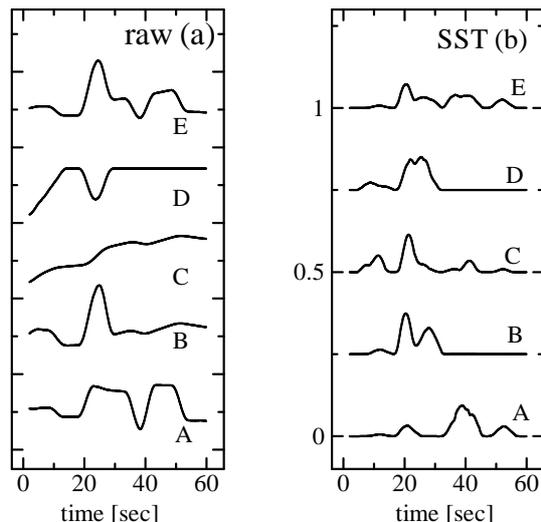


図4: 図3に対して、式(5)により平滑化をかけた結果($\sigma = 1.5$ 秒)。(a)元の時系列データ。(b)SSTを施された時系列。

に対する任意性が残る(したがって相対位置のみが意味を持つ)。

図5(a)によれば、BとEの距離が近いことがわかる。図3(a)を参照すると確かに両者には類似性があることがわかる。同様に、CとDも比較的近い。全体的に右肩上がりという傾向において両者は類似しており、他の変数に比べれば両者が近い位置に来るのは理解できる。

一方、図5(b)では、BとDが特に近接していることがわかる。これは、元の時系列図3(a)では似ても似つかなかったものであり、非常に興味深い結果といえる。この結果は、時系列同士はそれ自体では似ていなくても、変化のタイミングにおいて同期性が存在することに由来する。同様のことはCとEにも言える⁴。

実際のところ、この機械システムのデータは自動車を模したものであり、Bはエンジン回転数、Dはギヤ位置を表す。すなわちこの結果は「シフトダウンに伴いエンジン回転数が一時的に上昇した」という事実を反映するものになっている。事前知識なしにこのような因果律の一端が得られたことは注目に値する。先にも述べたように、SSTは元の時系列の表面的な振舞いの陰に隠れた因果律の痕跡を掘り起こし、ヘテロな変数群を共通の土俵で比較する有力な手段であることがわかる。

6 おわりに

特異スペクトル解析法を、時系列変数に対する非線形変換として再構成することにより、時系列群に隠された

⁴なお、図5(a)と(b)では、元の距離行列の定義が異なるため、データ点同士の距離の相対比較はできない。

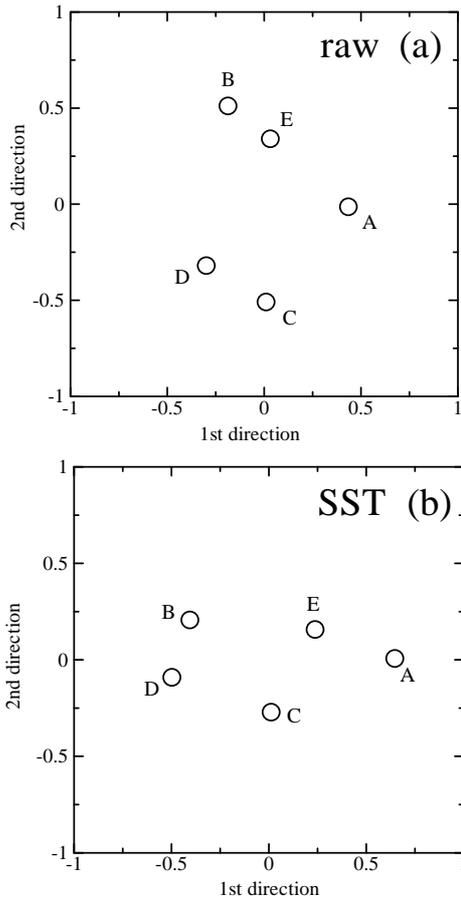


図 5: SST 前後での MDS プロットの比較。(a) 標準化変換した元時系列から L_2 距離で距離行列を構成した場合、(b) SST された時系列から Hellinger 距離で距離行列を構成した場合。

構造を発見する手法を新しく提案した。この変換は、時系列を変化度の時系列に直すことで、表面的な振る舞いの相違を剥ぎ取って、因果関係に基づく変数のクラスタリングを可能にする。

特異スペクトル変換は、ハンケル行列の特異値分解というノンパラメトリックな手法に基づく。この手法は、自己回帰モデルや混合正規分布モデルと異なり、異種混合の時系列セットに対して、柔軟に対応できる。

今回は変数個々の対に関してその関連性を述べたにとどまるが、それを一般化して、システム同士の比較の手段と考えることもできる。そのような視点から、我々は現在、この手法の実問題の応用を精力的に進めている。詳細な報告は別途行う予定である。

SST がデータマイニングの分野に提起する理論的課題は広範である。今回の手法に限って言えば、高速かつ逐次的な特異値分解算法の検討が課題として残る。また、今回は式 (1) の基準でパターン抽出を行ったが、これに関しては一般化しないし置換の余地が十分にある。距離行

列ないし類似度の行列の比較・分類手法を考えることは要するに、グラフという半構造データからの知識発見の問題に他ならない。最近、グラフマイニングの適用分野は動的システムにも拡張されつつあり [7]、その観点からも研究の発展が望まれる。

謝辞

本稿をまとめるにあたり、猪口明博氏および Nat Mills 氏との議論は有益であった。謝意を表する。

参考文献

- [1] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [2] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI-94 Workshop on Knowledge Discovery in Databases*. AAAI, 1994.
- [3] G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. In *Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1998.
- [4] C. Ding. Spectral clustering. In *Tutorial Notes of The Twenty-First International Conference on Machine Learning*, 2004.
- [5] M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40:1–41, 2002.
- [6] S. Hirano and S. Tsumoto. Mining similar temporal patterns in long time-series data and its application to medicine. In *Proceedings of 2002 IEEE International Conference on Data Mining (ICDM 2002)*, pp. 219–226. IEEE, 2002.
- [7] T. Idé and H. Kashima. Eigenspace-based anomaly detection in computer systems. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

- [8] E. Keogh, J. Lin, and W. Truppel. Clustering of time series subsequences is meaningless: Implications for previous and future research. In *IEEE International Conference on Data Mining (ICDM 2003)*. IEEE, 2003.
- [9] E. Keogh and M. Pazzani. Scaling up dynamic time warping for data mining applications. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 285–289. ACM, 2000.
- [10] K. Mardia. *Multivariate Analysis*. Academic Press, 1980.
- [11] V. Moskvina and A. Zhigljavsky. An algorithm based on singular spectrum analysis for change-point detection. *Communications in Statistics—Simulation and Computation*, 2003.
- [12] G. Strang. *Linear Algebra and its Applications*. Academic Press, 1976.
- [13] K. Yamanishi and J. Takeuchi. A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 676–681, 2002.
- [14] 尾崎統, 北川源四郎. 時系列解析の方法. 朝倉書店, 1998.
- [15] 福田剛志, 森本康彦, 徳山豪. データマイニング. 共立出版, 2001.
- [16] 柳井晴夫, 竹内啓. 射影行列・一般逆行列・特異値分解. 東京大学出版会, 1983.

A 補遺：MDSの古典解

A.1 中心化内積行列

まず距離行列の一般的性質について考えておく。今、時間点が p 個あるとして、 i 番目の時系列を $\mathbf{x}^{(i)}$ のように、 p 次元のベクトルとして表す ($i = 1, 2, \dots, h$)。距離行列を D と表すと、これは $h \times h$ 次元の行列となる。今、 D から、新たな行列 A を

$$A_{i,j} = -\frac{1}{2}D_{i,j}^2$$

によって作る。と、 $h \times h$ 行列 $B = HAH$ の要素は、

$$B_{i,j} = (\mathbf{x}^{(i)} - \bar{\mathbf{x}})^T (\mathbf{x}^{(j)} - \bar{\mathbf{x}}) \quad (6)$$

を満たすことを (多少の計算の後に) 示せる。 B は中心化内積行列 (centered inner product matrix) と呼ばれる。ここで $\bar{\mathbf{x}}$ は $\{\mathbf{x}^{(i)} | i = 1, \dots, h\}$ の平均ベクトルで、 H は h 次元の中心化行列 (centering matrix)

$$H = I - \frac{1}{h} \mathbf{1} \mathbf{1}^T$$

である。 h 次元単位行列を I 、要素がすべて 1 である $h \times h$ 行列を $\mathbf{1}$ と表した。

A.2 スペクトル定理の利用

式 (6) は要するに、行列 B が、中心化 (平均ベクトルを引き去ること) は別にして、元座標の 2 乗と関係付けられることを述べている。従って、距離行列 D から座標を復元するには、行列 B の「平方根」を見出せばよいことになる。実際、 h 個の点の元座標が、 $\{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(h)}\}$ のように与えられる時、行列 $R = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(h)}]$ に対して、

$$[R^T R]_{i,j} = \mathbf{y}^{(i)T} \mathbf{y}^{(j)} \quad (7)$$

が成り立つことが直ちにわかる。従って、何らかの方法で内積行列の「平方根」を求めれば、重心座標の任意性は別にして、元の座標を復元できることになる。

幸いにして、 B の固有値分解を使えばその「平方根」は容易に求められる。 B の固有ベクトルを、固有値の大きい順に q 個、列ベクトルとして並べた行列を $Q = [b_1, b_2, \dots, b_q]$ とおく。これは $h \times q$ 行列である。固有ベクトルのノルムは固有値の平方根に規格化しておく。すると、スペクトル展開 [12] における最初の q 項を取る近似において、行列 B は

$$B \simeq \sum_{i=1}^q b_i b_i^T = QQ^T \quad (8)$$

のように表せる。これはちょうど B の「平方根」が求まったことを意味する。式 (7) と式 (8) を見比べると、行列 Q の第 i 番目の行ベクトルが、 i 番目のデータ点の復元された座標に対応していることがわかる。

以上まとめると、距離行列 D から q 次元で座標を復元するには、まず、中心化内積行列を $B = HAH$ のように作り、固有値の大きい順に、その q 本の固有ベクトルを求めさえすればよい。そのノルムは固有値の平方根に規格化する。求められた固有ベクトルを列ベクトルとして並べた行列の、 i 番目の行ベクトルが、 i 番目のデータ点の復元された座標に対応する。