1031060322 松本　亮介

行列の圧縮による
変化点検出の高速化（理論）

0 はじめに

※この資料は，参考文献を元に，参考文献の概略を作成したものです．この研究を私が実際に行ったわけではありませんので注意してください．

1 まえがき

近年の多くの分野では，大量のデータを容易に入手できるようになり，そこからの知識発見に取り組むようになってきた．しかし，時系列データの変化点を検出する際に，古典的な手法では実用性にやや限界がある．それは，ある程度「おとなしい」振る舞いをする入力が，陰に陽に想定されているからである． 変化点検出については，多くの場合，不完全な事前知識に基づいて，しかも非定常な時系列を扱う必要がある．

このような，非定常な時系列の変化点を扱うための手法として，特異スペクトル変換（singular spectrum transformation; SST）と呼ばれる手法がある．SST は，ノンパラメトリックな変化点検出手法である．SST は，各時刻において，過去側と未来側の部分系列からの特徴抽出を行い，抽出された特徴ベクトル同士の食い違いをもって変化度スコアとする．SST は，特定の確率モデルを仮定しないので，入力時系列の多様性に比較的頑強であり，局所解の心配もない．

しかし，SST では，特徴ベクトル抽出に，特異値分解（singular value decomposition; SVD）を使うため，非常に多くの計算量になってしまい，それは，多くの利点を消し去るほどの欠点であった．
本稿では，SST における計算量を少なくし，高速化する手法について考える．変化点検出問題を，特徴ベクトル同士のカーネル(内積)の計算問題に帰着させ，そのカーネルを注意深い近似手法によって高速化する．

2 高速化計算手法

2.1 特異スペクトル変換

特異スペクトル変換について述べる．等間隔の時刻で観測された時系列データ $\{x_t \in \Re \mid t = 1,2,\ldots\}$ に対して，長さ w の部分系列を，$\Re^w$ の列ベクトルとして，

$$s(t) \equiv \left(x_{t-w+1}, \ldots, x_{t-1}, x_t\right)^{\mathrm{T}}$$

と定義する．
ある時刻 t において，部分系列を n 本並べた行列 $\mathrm{H}_1, \mathrm{H}_2$ を，

$$\begin{aligned} \mathrm{H}_1(t) &\equiv \left[s(t-n), \ldots, s(t-2), s(t-1)\right] \\ \mathrm{H}_2(t) &\equiv \left[s(t-n+\gamma), \ldots, s(t-1+\gamma)\right] \end{aligned}$$

と定義する．経験的に，$n = w$ とし，γ は $n/2$ とする．

SST ではこれらの行列から，特異抽出を SDV により行う．H_1 において特異値の大きい順に左特異ベクトルを r 個求める．それを，$U \equiv \left[u^{(1)}, u^{(2)}, \ldots, u^{(r)}\right]$ とする．
また，履歴空間 $\Pi_r \equiv \mathrm{span}\left\{u^{(1)}, u^{(2)}, \ldots, u^{(r)}\right\}$ とする．H_2 から最大左特異ベクトル μ を求める．
これより，t における変化度は，

$$1 - \mu^{\mathrm{T}} U^{\mathrm{T}} \mu / \left\| U^{\mathrm{T}} \mu \right\| \tag{1}$$

となり，第2項は超平面 Π_r と μ との余弦であり，この角度によって変化の度合いを見ている．

2.2 変化度式の変換

(1)を変更して，

$$z = 1 - \sum_{i=1}^{r} K(i, \mu)^2 \tag{2}$$

ただし，K は

$$K(i, \mu) \equiv \mu^{\mathrm{T}} u^{(i)} \tag{3}$$

で定義されるカーネル（内積）である．(2)はμで張られる空間と履歴空間の距離の 2 乗という意味を持つ．これらの計算を高速化するためには，いかに効率よく$K(i,\mu)$を計算するかである．

求めたいものは，内積であり，特異ベクトルではないので，特異ベクトルの計算をうまく省略して，内積を計算することを考える．また，必要な特異ベクトルの個数は履歴行列の次元よりはるかに小さいため，全ての特異ベクトルを求めることを避けるための計算手法を考える．

2.3 Krylov 部分空間の導入

H_1の最大左特異ベクトル$u^{(1)}$は，

$$u^{(1)} = \arg\max_u R(u) \tag{4}$$

ただし，

$$R(u) = \frac{u^{\mathrm{T}}\rho u}{u^{\mathrm{T}}u} \tag{5}$$

となる．この$R(u)$はレイリー商と呼ばれる量である．wより小さい任意の正整数sに対し，$s+1$次元の近似空間がs次元の近似空間よりも濃く最大固有状態を含むような，近似空間を作成することを考える．このためには，$\mathrm{span}\{\mu\}$に基底を付け加えて$\mathrm{span}\{\mu,\Delta\}$のような 2 次元空間を構成したとき，それが$R$の最急勾配方向を含む必要がある．

Rの勾配は，

$$\left.\frac{d}{du}R(u)\right|_{u=\mu} = \frac{-2}{u^{\mathrm{T}}u}\left[R(\mu)\mu - \rho\mu\right] \tag{6}$$

となる．これより，Δを$\rho\mu$と選べば，$\mathrm{span}\{\mu,\rho\mu\}$は，上記最急勾配方向を含むことがわかる．

この論法を続けていけば，k次元部分空間，

$$K_k(\mu,\rho) \equiv \mathrm{span}\left\{\mu,\rho\mu,\ldots,\rho^{k-1}\mu\right\} \tag{7}$$

は，Rの最大化の観点から最善のk次元部分空間であることがわかる．つまり，履歴行列の列空間（w次元）の部分空間として，μをひとつの基底として含む制約のもとで，ρの最大固有状態を最も濃く含むものが$K_k(\mu,\rho)$と考えられる．

2.4 固有値問題の圧縮

$K_k(\mu,\rho)$を張る正規直交基底を$\{q_1,\ldots,q_k\}$と表し，これらの列ベクトルとして持つ$w\times k$行列を，

$$\mathrm{Q} \equiv \left[q_1,\ldots,q_k\right] \tag{8}$$

と置く．Krylov 行列$\left[\mu,\rho\mu,\ldots,\rho^{k-1}\mu\right]$の QR 分解で求めることにすれば，QR 分解の一意性により，Q は一意的に求まる．

これより，$\rho u=\lambda u$は，$K_k(\mu,\rho)$へのuの射影$x\equiv \mathrm{Q}^{\mathrm{T}}u$を用いることで，

$$\mathrm{Q}^{\mathrm{T}}\rho\,\mathrm{Q}\,x = \lambda x \tag{9}$$

に近似的に変換される．これより，$k\times k$行列$\mathrm{Q}^{\mathrm{T}}\rho\mathrm{Q}$の対角化問題に帰着される．さらに，Krylov 行列と QR 分解より，

$$\mathrm{T} \equiv \mathrm{Q}^{\mathrm{T}}\rho\mathrm{Q} = \text{（対称 3 重対角行列）} \tag{10}$$

となっていることを示せる．これにより，$w\times w$行列ρに含まれていた情報が T の中の，わずか$2k-1$個の要素に圧縮されたといえる．対角要素$\alpha_1,\ldots,\alpha_k$と副対角要素$\beta_1,\ldots,\beta_{k-1}$は，Lanczos 反復といわれる方法で簡単に求めることができる．

2.5 カーネルの陰な計算

3 重対角行列の固有値問題は数値的に効率よくとくことができるとされている．(9)で求めた固有ベクトルを$\left\{x^{(1)},\ldots,x^{(k)}\right\}$とすると，$H_1$の固有ベクトルは，

$$u^{(\alpha)} = \sum_{i=1}^{k} q_i x_i^{(\alpha)} \tag{11}$$

ただし，$x_i^{(\alpha)}$はk次元ベクトル$x^{(\alpha)}$の第i成分を表す．ここで，$q_1=\mu$かつ$q_i^{\mathrm{T}}q_j=\delta_{i,j}$より，カーネルは，

$$K(\alpha,\mu) = x_1^{(\alpha)} \tag{12}$$

となる．これより，H_1の特異ベクトルを求める必要はなく，$k\times k$行列$\mathrm{Q}^{\mathrm{T}}\rho\mathrm{Q}$の対角化問題を求め，求めた固有ベクトルの第 1 成分を取れば，それより変過度を，

$$z = 1-\sum_{i=1}^{r}\left[x_1^{(i)}\right]^2 \tag{13}$$

と計算することができる．
これを，Implicit kernel 法と呼んでおく．

2.6 反復法による計算手法

カーネルを計算するには，μ（H_2 の最大左特異ベクトル）を求める必要がある．ここでは，べき乗法などの反復法に Zha-Simon 流のオンライン更新の考えを入れた一般的な枠組みとして，以下のように考える．
求めるには，a を正規化されたランダムベクトル $\widetilde{a}$ に初期化する．そして各 t において以下を行う．

1, $\mathrm{H}_2^{\mathrm{T}}\mathrm{H}_2$ に対する反復法を，a を初期ベクトルとして行い，最大固有ベクトルとして μ を求める．
2, $a = \mu + \varepsilon\widetilde{a}$ とおき，正規化する．

反復法には，べき乗法，EM-PCA，Lanczos3 重対角化と QR 反復の組み合わせなどがある．

2.7 パラメータの選択

パラメータの選択としては，w に関しては，「1 秒間以下で生じる細かい振動は無視する」，など，捉えたい変化の時間スケールから決めることができる．R に関しては，部分系列のパワースペクトル解析から（ほとんど w に依存せずに），相当広い範囲の時系列データで，数個程度でかまわないことがわかっている．明確な周期性のない時系列データについては，w によらず 3 から 4 程度，振動的な成分を含む時系列データについては，それよりも若干増やして 5 前後でよい．さらに，$K_k(\mu,\rho)$ の次元 k を与える必要があるが，Krylov 部分空間が，最大固有状態に加えて最小固有状態についてもよい近似空間になっていることから考えて，

$$k = \begin{cases} 2r & r \in \mathrm{even} \\ 2r-1 & r \in \mathrm{odd} \end{cases} \tag{14}$$

のように選ぶのが合理的である．これより，事前に入力すべきパラメータは，事実上 w のみとなる．この単純さが SST のひとつの利点でもある．

2.7 パラメータの選択

最後に計算手法を簡単にまとめておく．

まず，H_2 に関して，$\mathrm{H}_2\mathrm{H}_2^{\mathrm{T}}$ に対する反復法で，最大固有値をもとめ，最大固有ベクトル μ を求める．
次に，H_1 を Krylov 部分空間近似で圧縮し，3 重対角化行列 T に変換する．ここで，μ により，μ に無関係な成分は自動的に軽視するようにする．そして，μ は次の時間の行列における反復法の初期値とする．
そして，圧縮された行列 T を固有値分解して，$\{x^{(1)},\ldots,x^{(k)}\}$ を得る．最後に，それを利用して，変化度 z を求める．

2 種類の特異ベクトル同士の比較は陰になされていて，面倒な計算が全て，圧縮された空間内で行われている．

3 まとめ

実験の結果，かなり効率よく計算することができた．変化度計算をカーネルの計算に帰着させ，Lanczos 法により有効な情報を濃縮した上で，陰にカーネルを計算する．

一般に次元が数百になると，H_1 のような密行列の複数の特異ベクトルを反復法で計算するのは，困難になる．求めたい固有ベクトルが増えるにつれ，大幅に計算量が増加し，計算速度が大きく低下する．実際にランダムベクトルから，陽に特徴ベクトルを求めると，特異ベクトルの偽縮退が現れる．これがおきると，基底がばらばらになり，変化度が不安定になる．

提案手法では，密行列の情報を圧縮することで，数値的に扱いやすい低次元空間で，全ての計算を済ませている．変化度 z を求めるにあたって，履歴空間 Π_r が直接必要ではないため，それをうまく利用して危険の多い $\{u^{(i)}\}$ の計算を回避することで，高速に SST を計算することに成功した．

参考文献

[1]　井手剛　“行列の圧縮による変化点検出の高速化”