

部分時系列クラスタリングの理論的基礎

Theoretical basis for subsequence time-series clustering

井手剛

Tsuyoshi Idé

IBM 東京基礎研究所

IBM Research, Tokyo Research Laboratory

It is known that the k -means cluster centers in sliding-window-based subsequence time-series clustering (STSC) become sinusoidal pseudo-patterns under a fairly general condition. We theoretically discuss the origin of this counter-intuitive effect. Based on a spectral clustering formulation, we explicitly show how the STSC cluster centers become sinusoids.

1. はじめに

部分時系列クラスタリング (subsequence time-series clustering; STSC) は、時系列データからのパターン抽出方法としてもっとも有名なもののひとつである。これは、1本の時系列から部分系列を多数生成し、それらを独立なベクトルとしてクラスタリングを実行することで、クラスタ中心を「パターン」として取り出す手法である。部分系列の生成には、固定窓サイズ w の滑走窓方式が使われるのが典型的である (図 1 (a))。本論文でも専ら固定窓幅の滑走窓方式の STSC を考える。

今まで、この方法に基づいて数多くの「発見」がなされてきた (例えば [2])。しかし 2003 年に、滑走窓方式の k 平均 STSC がパターン抽出手法としては無意味であるとの衝撃的な報告 [6] がなされて以降、この分野の研究に沈滞感が否めない。

Keogh ら [6] が実験的に指摘したのは、かなり広い条件下、 k 平均のクラスタ中心が、入力時系列とほぼ無関係に正弦波パターンとなっているという事実であった。読者の便宜のため、彼らの結果を図 2 (a) に再現した。データはいわゆる CBF データ [5] で、図 1 (b) に示すような長さ 128 のインスタンスをそれぞれ 30 個連結して、図 1 (a) に例示したような 1 本の長い時系列を作る。例えば Cylinder の立ち上がりの位置やその長さは、異なるインスタンスにわたりランダムに変化する。他も同様である。これに対して k 平均 STSC を実行し、クラスタ中心を計算したのが図 2 (a) である。図からわかるように、元の CBF パターンとは似ても似つかぬ波長 w の正弦波が得られている。

STSC におけるこの正弦波効果について、現象の劇的さの割には、理論的にわかっていることは少ない。その発現機構の理論的説明は、データマイニングおよび機械学習における主要な未解決問題のひとつであると言ってよい。本論文の目的は、STSC の理論的な基礎を吟味し、正弦波効果がどのような条件下でいかに生じるのかをあらわに明らかにすることである。われわれの知る限り、本論文は、正弦波効果に完全な理論的説明を与えた初めての仕事である。

ところで、面白いことに、正弦波効果は、部分系列を列ベクトルとして持つ行列 H の特異ベクトルにおいても生じる。図 2 (b) は、連結 CBF データに対して最初の (固有値の大きい) 3 つの特異ベクトルを図示したものである。第 1、第 2 特異ベクトル ($u^{(1)}$ と $u^{(2)}$) は、 k 平均のクラスタ中心とよく似た波長 w の正弦波となっている。

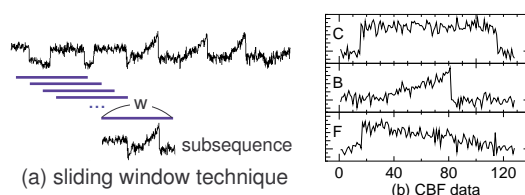


図 1: (a) 滑走窓方式の STSC (データは連結 CBF データの一部)。 (b) Cylinder (C)、Bell (B)、Funnel (F) データのインスタンスの例。

この類似は、 k 平均と H の特異値分解 (SVD) が、ある共通の数学的構造を持っており、そしてその構造こそが正弦波効果を生み出す原因となっていることを示唆する。この予想に基づいて、われわれは、次節以降において、まず k 平均 STSC におけるパターン発見問題を spectral クラスタリング [10, 7, 3] の形に書き換える。そして、問題の持つ並進対称性に注目することで、STSC のクラスタ中心が正弦波となることをあらわに示す。

ここで、正弦波効果とわれわれが呼ぶ実験事実について改めてまとめておこう。CBF データについて言えば、

実験事実 1 k 平均 STSC のクラスタ中心は波長 w の正弦波でよく近似される。 k 個の正弦波の絶対位相はランダムであるが、相対位相は $2\pi/k$ の整数倍となる。

実験事実 2 部分系列行列 H の左特異ベクトルは正弦波でよく近似される。最初のいくつかは波長 w を持つ。

以下、第 2 節で、時系列マイニングのための新しいモデルを導入する。仮想的に 1 次元周期格子を導入し、その上で線形演算子を定義することにより、STSC の裏に潜む対称性をあらわに解析できるようになる。第 3 節では、 k 平均のクラスタ中心が、ある固有値問題を解くことで直接得られることを示す。第 4 節では、クラスタ中心が正弦波となることをあらわに示す。第 5 節で実験結果を補足した後、最後にまとめを行う。

2. 基礎事項

2.1 時系列解析のための 1 次元格子モデル

時系列 Γ を、 n 個の実数 x_1, x_2, \dots, x_n の順序つき集合として定義する。 Γ に対して、長さ $w \leq n$ の部分系列 s_p を、 $(x_p, x_{p+1}, \dots, x_{p+w-1})$ として定義する。 s_p を w 次元のベクトル

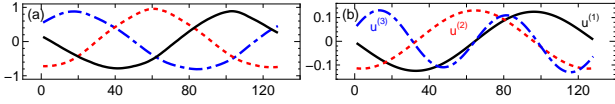


図 2: (a) k 平均クラスタ中心 ($k = 3, w = 128$). (b) 最初の 3 つの左特異ベクトル ($w = 128$).

ルとして見る時には、 s_p と書くことにする。STSC においては s_p は独立したデータ要素として扱われる。

n 格子点 (もしくはサイト) をもつ 1 次元格子を考え、各サイトを $l \in \{1, 2, \dots, n\}$ でラベル付ける。時系列 Γ は、格子上的「状態」として理解することができる。状態を表すために、正規直交基底 $\{|1\rangle, |2\rangle, \dots, |n\rangle\}$ で張られる n 次元のヒルベルト空間 \mathcal{H}_0 を定義する。ここで格子上的状態を表すためにディラック記法 [8] を使った*1。

双対空間 $\mathcal{H}_0 = \{|1\rangle, \dots, |n\rangle\}$ を導入することにより、 $|l\rangle$ と $\langle l'|\langle l'|\rangle$ の間の内積を $\langle l|l'\rangle$ と表す。基底の正規直交性は、したがって、 $\langle l|l'\rangle = \delta_{l,l'}$ のように書ける ($\delta_{l,l'}$ はクロネッカーのデルタ)。基底の正規直交性から、任意の Γ は

$$|\Gamma\rangle = \sum_{l=1}^n x_l |l\rangle \quad (1)$$

のような線形結合として書ける。このような表示をサイト表示と呼んでおく。 $x_l = \langle l|\Gamma\rangle$ である。サイト表示では係数 x_l は実数であるが、一般に、ヒルベルト空間の元の内積について、

$$\langle \Gamma|l\rangle = \langle l|\Gamma\rangle^* \quad (2)$$

が成り立つと約束する (* は複素共役)。

2.2 \mathcal{H}_0 での線形演算子

\mathcal{H}_0 上でのある状態を別の状態に移す線形演算子の集まりを \mathcal{L} と表す。普通の数と演算子を区別するために、以下、線形演算子には $\hat{\cdot}$ をつける。定義から、 $\forall \hat{o} \in \mathcal{L}$ は、サイト表示において

$$\hat{o} = \sum_{l,l'=1}^n o_{l,l'} |l\rangle \langle l'| \quad (3)$$

のような表現をもつはずである。ここで、 $o_{l,l'}$ を \hat{o} の (l, l') 成分 (もしくは行列要素) と呼ぶ。定義から $o_{l,l'} = \langle l|\hat{o}|l'\rangle$ が成り立つ。

われわれの文脈では、並進演算子 $\hat{\tau}(l)$

$$\hat{\tau}(l) \equiv \sum_{l'=1}^n |l'+l\rangle \langle l'| \quad (4)$$

が特に重要である。このように、演算子の作用をあらわに表せる点がディラック記法のご利益である。 $\forall |\Gamma\rangle \in \mathcal{H}_0$ に対して、 $\hat{\tau}(l)|\Gamma\rangle$ の双対元を $\langle \Gamma|\hat{\tau}(l)^\dagger$ と表し、 $\hat{\tau}(l)^\dagger$ を $\hat{\tau}(l)$ のエルミート共役と呼ぶ。定義から、 $\hat{\tau}(l)^\dagger = \hat{\tau}(-l)$ が成り立つ。

以下、理論解析上、周期的境界条件、 $\forall l, |l+n\rangle = |l\rangle$ を仮定する。 $n \gg 1$ である限り、この人為的条件の効果は無視できるはずである。

*1 数学的にはディラック記法は伝統的なベクトル・行列記法と同等であるが、後でわかるように格子上での線形演算子の表現能力において格段に優れており、STSC の解析には最善の選択だと思われる。

2.3 離散フーリエ変換

$w \leq n$ に対して部分空間 $\mathcal{H} = \{|1\rangle, \dots, |w\rangle\} \subseteq \mathcal{H}_0$ を考える (\mathcal{H} では $w = n$ でない限り周期的境界条件は課さない)。ここで、サイト表示からフーリエ表示に表現を移す変換

$$|f_q\rangle = \frac{1}{\sqrt{w}} \sum_{l=1}^w e^{if_q(l-l_0)} |l\rangle; |l\rangle = \frac{1}{\sqrt{w}} \sum_{q \in \mathcal{D}_f} e^{-if_q(l-l_0)} |f_q\rangle$$

を定義する。ここで、 l_0 は任意の実数であり、 $f_q = 2\pi q/w$ である。添え字 q は w が奇の時は $\mathcal{D}_f = \{-\frac{w-1}{2}, \dots, 0, 1, \dots, \frac{w-1}{2}\}$ を走り、偶の時は $\{-\frac{w}{2} + 1, \dots, 0, 1, \dots, \frac{w}{2}\}$ を走る。定義から $\langle f_q|f_{q'}\rangle = \delta_{q,q'}$ は容易に示せる。したがって、 $\{|f_q\rangle\}$ は \mathcal{H} において完全系をなす。

任意の状態 $|\gamma\rangle \in \mathcal{H}$ に対し、離散フーリエ変換 (discrete Fourier transformation; DFT) を、

$$|\gamma\rangle = \sum_{q \in \mathcal{D}_f} |f_q\rangle \langle f_q|\gamma\rangle; \langle f_q|\gamma\rangle = \sum_{l=1}^w \langle f_q|l\rangle \langle l|\gamma\rangle \quad (5)$$

で定義する。ここで定義より、 $\langle f_q|l\rangle = \frac{1}{\sqrt{w}} e^{-if_q(l-l_0)}$ である。もし γ が実数値時系列の表現であれば、 $|l\rangle$ の重みは実であるから、次が成り立たねばならない。

$$\langle l|\gamma\rangle = \frac{1}{2\sqrt{w}} \sum_{q \in \mathcal{D}_f} [e^{if_q(l-l_0)} \langle f_q|\gamma\rangle + e^{-if_q(l-l_0)} \langle f_q|\gamma\rangle^*]. \quad (6)$$

3. 密度演算子による k 平均法の定式化

3.1 k 平均法の目的関数

状態の集まり $\{|s_q\rangle \in \mathcal{H} | q = 1, 2, \dots, n\}$ に対し、 k 平均クラスタリングを行うことを考える。 k 平均は 2 乗誤差を最小にするアルゴリズムであることがよく知られている [4]。容易にわかるように、われわれの記法では 2 乗誤差は

$$E = \sum_{p=1}^n \langle s_p|s_p\rangle - \sum_{j=1}^k |C_j| \langle m^{(j)}|m^{(j)}\rangle \quad (7)$$

のように書かれる。ここで、 C_j は j 番目のクラスターのメンバーを表し、 $|C_j|$ はそのクラスターに属するデータの数を表す。また、 C_j のセントロイドを $|m^{(j)}\rangle$ と書いた。

第 1 項はクラスタリング結果に無関係だから、第 2 項 (E_2 と書く) に注目しよう。セントロイドの定義 $|m^{(j)}\rangle = \frac{1}{|C_j|} \sum_{p \in C_j} |s_p\rangle$ を代入して整理することにより、

$$E_2 = - \sum_{j=1}^k \frac{1}{|C_j|} \sum_{p,r \in C_j} \langle s_p|s_r\rangle \quad (8)$$

となる。制約付きの和を取り除くために、ここで「割り当てベクトル」 $|u^{(j)}\rangle \in \mathcal{H}$ を導入する。その定義は、 $s_q \in C_j$ に対し $\langle s_q|u^{(j)}\rangle = 1/\sqrt{|C_j|}$ 、それ以外だとゼロ、である。その結果、

$$E_2 = - \sum_{j=1}^k \sum_{p,r=1}^n \langle u^{(j)}|s_p\rangle \langle s_p|s_r\rangle \langle s_r|u^{(j)}\rangle.$$

となる。この意味をさらにはっきりさせるために、線形演算子

$$\hat{\rho} = \sum_{p=1}^n |s_p\rangle \langle s_p|$$

を導入する。統計力学的イメージ [8] に基づいて、 $\hat{\rho}$ を密度演算子と呼んでおく。 $|s_p\rangle$ たちは、滑走窓のずらして得られるので、並進演算子を用いれば、

$$\hat{\rho} \doteq \sum_{l=1}^n \hat{\tau}(l)^\dagger |\Gamma\rangle \langle \Gamma| \hat{\tau}(l) \quad (9)$$

のようにも書ける。ただし“ \doteq ”は、「 $(\mathcal{H}_0$ ではなく) \mathcal{H} において表現された時に、左辺と右辺が同一の行列要素を持つ」ことを表す。結局、 $\hat{\rho}$ を使うと k 平均の目的関数は

$$E_2 = - \sum_{j=1}^k \langle u^{(j)} | \hat{\rho} | u^{(j)} \rangle \quad (10)$$

となる。なすべきなのは、 E_2 を最小化する状態 $\{|u^{(j)}\rangle\}$ を求めることである。

3.2 固有値方程式との関係

これまでのところ、 $|u^{(j)}\rangle$ は人為的に導入された「クラスタ割り当てベクトル」であった。元の定義から、下記が満たされることは容易にわかる。

$$\sum_{p=1}^n \langle u^{(i)} | s_p \rangle \langle s_p | u^{(j)} \rangle = \langle u^{(i)} | \hat{\rho} | u^{(j)} \rangle = \delta_{i,j} \quad (11)$$

ここで、ゼロか非ゼロかの二分法を緩和し、上記の条件を、割り当てベクトルに対する新しい制約条件式と考える。これは $|u^{(j)}\rangle$ に対する一般化された直交条件であるから、結局 k 平均クラスタリングは一般化固有値問題と同等ということになる。今の場合は、下記の固有値問題とも同等である。

$$\hat{\rho} |u^{(j)}\rangle = \lambda_j |u^{(j)}\rangle \quad \text{s.t.} \quad \langle u^{(i)} | u^{(j)} \rangle = \delta_{i,j}. \quad (12)$$

ここで λ_j は固有状態 $|u^{(j)}\rangle$ に対する固有値で、値が大きい順にラベル付けされているとする。サイト表示の基底を作用させて、上式は

$$\sum_{l'=1}^w \langle l | \hat{\rho} | l' \rangle \langle l' | u^{(j)} \rangle = \lambda_j \langle l | u^{(j)} \rangle$$

とも書ける。普通のベクトル・行列表記では、 $\langle l | \hat{\rho} | l' \rangle$ は HH^T の (l, l') 要素に対応する。ただし $\text{H} = [s_1, \dots, s_n]$ で (周期的境界条件に注意) $^\text{T}$ は転置を表す。 $\langle l | u^{(j)} \rangle$ を w 次元列ベクトル $u^{(j)}$ の j 番目の要素だと思えば、結局これは、 HH^T に対する固有値問題と同じである。あるいは、 H の左特異ベクトルを求めているのと同じである [9]。

3.3 クラスタ中心との関係

$|u^{(j)}\rangle$ の意味をさらに詳しく見よう。問題の緩和前は、セントロイドと割り当てベクトルの定義から、

$$|m^{(j)}\rangle \equiv \frac{1}{|C_j|} \sum_{p \in C_j} |s_p\rangle = \frac{1}{\sqrt{|C_j|}} \sum_{p=1}^n |s_p\rangle \langle s_p | u^{(j)} \rangle$$

が満たされていた。緩和後は、 $|u^{(j)}\rangle$ は $\hat{\rho} = \sum_p |s_p\rangle \langle s_p|$ の固有ベクトルなのだから、上式より直ちに以下を得る:

$$|m^{(j)}\rangle \propto |u^{(j)}\rangle. \quad (13)$$

これも含め、この節の結果を定理としてまとめておこう。

定理 1 $\hat{\rho}$ の固有状態は、 H の左特異ベクトルとしても計算できる。それらは k 平均法と同じ目的関数を最小化する。

定理 2 $\hat{\rho}$ の固有状態は k 平均のクラスタ中心に対応する。

これらの結果から、われわれは、 k 平均でのクラスタ中心の挙動を、固有値問題として調べることができる。しかし、両者の対応は必ずしも完璧ではないことに注意を要する。主たる相違は、クラスタ中心同士に直交条件が課せられるか否かにある。この場合にもっともらしい想定としては、最大固有状態 $|u^{(1)}\rangle$ は、主たる k 平均クラスタのよい近似になっているであろうというものである。他の固有状態に対しては、多かれ少なかれ $|u^{(1)}\rangle$ に対する直交条件により影響を受けるだろう。この辺については、後にちょっと触れる。

4. 正弦波解の導出

4.1 $\hat{\rho}$ のフーリエ表現

$\hat{\rho}$ の固有値問題をあらわに解くことで、 k 平均 STSC の解析解を求めてみよう。最初に $w = n$ の場合を考える。この時、 $\mathcal{H} (= \mathcal{H}_0)$ は周期的と見なせ、ゆえ、フーリエ状態 $|f_q\rangle$ は $\hat{\tau}(l)$ の固有状態になっている。すなわち、

$$\hat{\tau}(l) |f_q\rangle = \frac{1}{\sqrt{w}} \sum_{l'=1}^n e^{if_q(l'-l_0)} |l'+l\rangle = e^{-if_q l} |f_q\rangle. \quad (14)$$

ここで $f_q = 2\pi q/n$ に対して $e^{if_q n} = 1$ が成り立つことを用いた。式 (9) と (14) を使って、行列要素 $\langle f_q | \hat{\rho} | f_{q'} \rangle$ は、

$$\sum_{l=1}^n \langle f_q | \Gamma \rangle \langle \Gamma | f_{q'} \rangle e^{i(f_q - f_{q'})l} = n |\langle f_q | \Gamma \rangle|^2 \delta_{q,q'} \quad (15)$$

のように計算できる。すなわち、この時は、 $\hat{\rho}$ はフーリエ表示において対角的であり、このことは、フーリエ状態そのものが固有状態になっていることを意味する。どの f_q が選ばれるかは、フーリエ成分のパワー $|\langle f_q | \Gamma \rangle|^2$ に依存し、例えば、最大固有ベクトルは最大パワーを持つ f_q で生ずる。

$w < n$ の時は厳密解は一般には得られないが、摂動論的に固有状態を調べることができる。詳細は省くが、結果は、

$$\langle f_q | \hat{\rho} | f_{q'} \rangle \approx n |\langle f_q | \Gamma \rangle|^2 \delta_{q,q'} + \sum_{l=1}^n e^{i\Delta_{q',q} l} J_l(q, q') \quad (16)$$

のようになる。ここで $\Delta_{q',q} = f_{q'} - f_q$ である。 $J_l(q, q')$ の表式は略すが、もし Γ が支配的な $|f_q\rangle$ を持っていれば、非対角成分としての寄与は非常に小さいことを示せる。

4.2 固有ベクトルの形

ちょうど連結 CBF データのように、かなり乱れはありつつも時系列 Γ がほぼ周期的である場合を考えよう。この時、周期性から考えて、特定の $|f_q\rangle$ のパワーが大きい値を持つはずであり、上記の通り、固有値問題の解は $|f_q\rangle$ に非常に近いことが予想される。直感的には、滑走窓方式が、部分系列同士のコヒーレンス (整位相性) を保つ方式であるがゆえ、データのばらつきは位相相殺により大幅にならされてしまい、共通特徴としての $|f_q\rangle$ 成分のみが非常に強められる、と見ることもできよう。

いずれにしる、 $|\langle f_q | \Gamma \rangle|$ が支配的な場合、式 (6) により、固有ベクトルもしくはクラスタ中心は、

$$\langle l | u \rangle \propto \cos(f_q l + \phi) \quad (17)$$

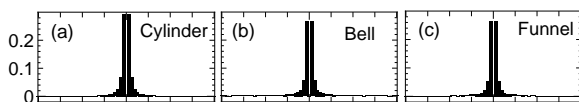


図 3: CBF データのパワースペクトル。それぞれ 30 インスタンスにわたる平均で、横軸は波数 ($-\pi/2$ から $\pi/2$)。

のような形になるはずである。 l_0 は任意であったから、ここで便宜上導入した ϕ もまた任意の位相因子である。

以上の結果を定理としてまとめておこう。

定理 3 ある $|q|$ が支配的な時、 H の最大左特異ベクトルは、波長 $w/|q|$ で、任意の位相因子を持つ正弦波でよく近似される。これは入力時系列の詳細にはよらない。

後で述べるように、定理 2 と併せて、これは実験結果 1 に理論的説明を与えている。また、証明は略すが、クラスタ中心の間の相対位相については下記の定理が成り立つ。

定理 4 時系列 Γ が支配的な $|f_q|$ を持つ時、 k 平均 STSC は、波数 f_q を持ち一様に位相が分布した正弦波のクラスタリング問題に帰着できる。 2 乗誤差基準の観点で最適なクラスタ中心は、互いに $2\pi/k$ の整数倍の位相差を持つ。

5. 実験結果

われわれの理論が CBF データにおける正弦波効果を説明することを示すために、まず、部分系列のパワースペクトルを計算した。結果を図 3 (a)-(c) に掲げる。それぞれ 30 個のインスタンスがあるが、示したものは、得られたスペクトルを総和平均したものである。インスタンスのランダムな変動にもかかわらず、圧倒的に $|q| = 1$ の成分が大きいことがわかる ($w = 128$)。連結前に各インスタンスは平均ゼロ、分散 1 に標準化されているので、 f_0 成分のパワーは当然 0 である。

k 平均 STSC および対応する SVD の結果は、冒頭図 2 に示してある。そこで得られた波長 w は、 $|q| = 1$ が支配的であることから理解できる (定理 2 および 3)。また、図 2 (a) では 3 つの正弦波は相対的に $2\pi/3$ だけ位相が異なっており、定理 4 は正確にその説明を与えている。

一方、図 2 (b) では、 $u^{(1)}$ と $u^{(2)}$ は位相 $\pi/2$ だけ隔たっており、 k 平均の結果と異なっている。また、 $u^{(3)}$ は $u^{(1)}$ と同位相だが、波長が $w/2$ の正弦波となっている。波長 w の範囲では互いに直交する波は 2 つしか作れないから、次の波長は必然的に $w/2$ である。すなわち $u^{(3)}$ の波長は直交条件の必然的帰結である。これは、二つの定式化において、クラスタ中心同士のいわば相互作用の仕方の違いを表す一つの典型的な例である。

なお、STSC における正弦波効果を、ランダムインスタンスの重ね合わせによる素朴な平滑化効果に帰す見方があるが (例えば [1])、その理解は正確とはいえない。紙幅の関係で紹介できないが、例えば、酔歩データの STSC においても、ある条件下で正弦波が発生する。また、パワースペクトルの形によっては、物理学でいう「うなり」現象すら起こる。これらはいずれも、単純な平滑化効果では説明しようもないものである。

6. まとめ

滑走窓方式の STSC についての理論解析を行い、正弦波効果を理論的に説明することに初めて成功した。道具立てとし

ては、 k 平均法を spectral クラスタリングの形に再定式化し、その解を、1 次元周期格子モデル上の並進演算子を使って調べた。その結果、DFT において支配的な $|f_q|$ があれば、結果として、同一の $|f_q|$ を持つ正弦波の集まりをクラスタリングしているのと同じであることがわかった。 k 平均と spectral クラスタリングでは、クラスタ中心の相互作用に相違があるが、正弦波形成機構においては本質的に変わりはない。

時系列データに対する有効なパターン抽出手法を考えるとこの上では、本論文の結果は非常に重要な出発点になりえる。コヒーレントな重ねあわせでクラスタ中心を表現している限り、正弦波効果を回避することは難しい。よりよい時系列パターン抽出方法についての詳しい議論は別稿に譲る。

参考文献

- [1] J. Chen. Making subsequence time series clustering meaningful. In *Proc. IEEE Intl. Conf. on Data Mining*, pages 114–121, 2005.
- [2] G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. In *Proc. the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1998.
- [3] I. S. Dhillon, Y. Guan, and B. Kulis. Kernel k-means, spectral clustering and normalized cuts. In *Proc. Tenth ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 551–556, 2004.
- [4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd ed.)*. Wiley Interscience, 2000.
- [5] E. Keogh and T. Folias. The UCR time series data mining archive [<http://www.cs.ucr.edu/~eamonn/TSDMA/index.html>]. 2002.
- [6] E. Keogh, J. Lin, and W. Truppel. Clustering of time series subsequences is meaningless: Implications for previous and future research. In *Proc. IEEE Intl. Conf. on Data Mining*, pages 115–122. IEEE, 2003.
- [7] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems, 14*, pages 849–856, 2001.
- [8] J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 2nd Ed., 1994.
- [9] G. Strang. *Linear Algebra and its Applications*. Academic Press, 1976.
- [10] H. Zha, C. Ding, M. Gu, X. He, and H. Simon. Spectral relaxation for k-means clustering. In *Advances in Neural Information Processing Systems, 14*, pages 1057–1064, 2001.