

DTW 距離を用いた時系列データのベクトル空間への埋込

水原 悠子[†] 林 朗[†] 末松 伸朗[†]

Embedding Time Series Data Using Dynamic Time Warping Distances

Yuko MIZUHARA[†], Akira HAYASHI[†], and Nobuo SUEMATSU[†]

あらまし 動的時間伸縮 (DTW) により得られたデータ間距離をもとに、時系列データをベクトル空間へ埋込み、埋込空間にて分類するアプローチを提案する。ラベル付データとラベルなしデータが最初に与えられる学習問題設定のもとで、埋込手法の候補として、MDS によるユークリッド空間への埋込、擬似ユークリッド空間への埋込、及びラブラシアン固有マップ法による埋込の 3 手法を考える。DTW 距離の性質と埋込手法の適合性に関する考察、及び分類実験により、ラブラシアン固有マップ法による埋込が 3 手法の中で最も高い分類精度につながる事が分かった。また、ラブラシアン固有マップ法を用いた提案アプローチは k -近傍法より高い分類精度を実験で示した。

キーワード 機械学習, 時系列, パターン認識, DTW, Kernel PCA

1. ま え が き

1.1 時系列データの分類問題

情報技術の発展とともに、ジェスチャー認識、ビデオ検索、オンライン手書き文字認識など時系列データ認識の重要性が高まっている。本研究では次のような時系列データの 2 クラス分類学習問題を考える。

n 個の時系列データからなる集合 $\mathcal{X} = \{X_1, \dots, X_n\}$ が与えられたとする。ここに、 X_i ($1 \leq i \leq n$) は特徴ベクトルから構成される長さ l_i の系列: $X_i = (x_1^i, \dots, x_{l_i}^i)$ である。時系列データのうち、最初から s 個、 $\{X_i \mid 1 \leq i \leq s\}$ はラベル付データであり、クラスラベル $y_i \in \{-1, +1\}$ が与えられている。このとき、ラベルなしデータ $\{X_i \mid s+1 \leq i \leq n\}$ のクラスを推定せよ。

時系列データは、データにより長さが異なるなど、固定次元のベクトルデータとは大きく異なる特性をもち、その扱いは困難である。

時系列データの分類手法として、音声認識の分野で成功した隠れマルコフモデル (Hidden Markov Model: HMM) [14], [24] など、生成モデルを用いた手法があ

る。正しくモデルが推定された場合には精度が高い。しかし、多くの訓練データを必要とし、また、モデル学習の際、局所解に陥りやすいといった問題点がある [25]。

一方、距離ベースの手法として、 k -近傍法など、動的時間伸縮 (dynamic time warping: DTW) [14] により得られる時系列データ間距離を用いて分類を行う手法がある [6]。 k -近傍法は生成モデルとは異なり、確率密度関数の形を想定しなくてよく、また、複雑な識別境界を表現することができる。しかし、一般にノイズに敏感であるという問題点がある [22]。

1.2 提案アプローチ

本研究では、大量の訓練データを必要としない距離ベースの手法に着目する。提案アプローチを以下に示す。

- (1) DTW により時系列データ間距離を計算する。
- (2) (何らかの意味で) DTW 距離を保存するような写像 Φ により、時系列データをベクトル空間 (特徴空間) \mathcal{F} へ写像する。

$$\Phi: \mathcal{X} \rightarrow \mathcal{F}$$

$$X_i \mapsto \Phi(X_i)$$

次に $\Phi(X_i)$ を低次元部分空間へ射影して、 $\tilde{\Phi}(X_i)$ を得る。

- (3) \mathcal{F} にて、ラベル付データ $\{(\tilde{\Phi}(X_i), y_i) \mid 1 \leq$

[†] 広島市立大学大学院情報科学研究科, 広島市 Graduate School of Information Sciences, Hiroshima City University, 3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima-shi, 731-3194 Japan

$i \leq s\}$ を用いて、分類器を学習する。

(4) 分類器を用いて、ラベルなしデータ $\{\tilde{\Phi}(X_i) \mid s+1 \leq i \leq n\}$ を分類する。

1.3 ベクトル空間への埋込手法

本研究では時系列データのベクトル空間への埋込手法、すなわち写像 Φ として、三つの手法を考える。手法1は多次元尺度法 (multidimensional scaling: MDS) [21] であり、ユークリッド空間への埋込を行う。手法2は手法1の拡張であり、擬似ユークリッド空間への埋込を行う [7], [8], [13]。手法3は多様体学習手法 [2], [3] を用いて、ユークリッド空間への埋込を行う。

三つの埋込手法を検討する理由は以下のとおりである。MDS (手法1) は、データ間距離に基づく埋込手法としてよく用いられている。手法2で擬似ユークリッド空間への埋込を行うのは、DTW 距離が (ユークリッド) 距離の三角不等式を満たさないためである。手法3ではいくつかある多様体学習手法のうち、非ベクトルデータにも適用できるものを選んだ。一般に、多様体学習は非線形次元縮小手法として近年注目を集めており、多様体上の測地線距離を保存するような低次元空間への埋込を行う [15], [20]。

本研究では、3手法を分析する上で、主成分分析 (principal component analysis: PCA) の拡張であるカーネル PCA [16] の理論的枠組みを用いる。その理由は以下のとおりである。(1) 三つの手法をカーネル PCA と関連づけることにより、各手法の位置付け、比較考察が容易になる。(2) \mathcal{X} に含まれない新規時系列の埋込 [4]、あるいは、新たなカーネルを用いた埋込など、提案アプローチの拡張が期待できる。

1.4 関連研究

Shimodaira ら [18] はカーネル関数に非線形時間伸縮の機能を導入した SVM を提案して、音声の認識を行い、訓練データ数が少ない場合、HMM よりも高い分類精度を得ている。Bahlmann ら [1] はガウシアンカーネルの距離の項に DTW 距離を代入した GDTW カーネルを提案し、オンライン手書き文字認識にて HMM に比肩する分類精度を得ている。しかし、いずれもカーネル行列の半正定値性が証明できず、特徴空間 (ヒルベルト空間) の存在を保証することができないという問題点がある。

Graepel ら [8] はデータ間の類似度に基づき、データを擬似ユークリッド空間に埋め込み、SVM を用いて分類する手法を提案し、猫の脳皮質やタンパク質

データの分類実験で、 k -近傍法に比較して良好な結果を得ている。Pekalska ら [13] も同様な手法を提案し、オフライン手書き数字分類、2値画像からの形状分類問題にて良好な結果を報告している。しかしいずれも時系列データの分類は扱っていない。

Belkin ら [2], [3] はラプラシアン固有マップ法を提案し、類似度に基づきデータを低次元ベクトル空間へ埋め込み、手書き数字の分類実験で、 k 近傍法に比較して良好な結果を得ている。しかし時系列データの分類は扱っていない。

1.5 論文の構成

以下、DTW (2.)、及びカーネル PCA (3.) について説明した後、4. でベクトル空間への埋込手法、5. で埋込前後のデータ間距離、6. で分類器の学習と分類について述べる。7. では提案手法を評価するために行った実験について説明し、8. でまとめる。

2. DTW

DTW は二つの特徴ベクトル時系列に対して時間伸縮を許して可能なすべての対応を評価し、その中で距離最小、すなわち類似度最大となる対応付けを見出すものである [14]。本研究で用いた時系列間 DTW 距離 $\{d^2(X_i, X_j) \mid 1 \leq i, j \leq n\}$ は、以下の漸化式で計算した。

(1) 初期値: $g(0, 0) = 0$

(2) 繰返し: $1 \leq t_i \leq l_i, 1 \leq t_j \leq l_j$ において

$$g(t_i, t_j) = \min \begin{cases} g(t_i - 1, t_j) + \|\mathbf{x}_{t_i}^i - \mathbf{x}_{t_j}^j\|^2 \\ g(t_i - 1, t_j - 1) + 2\|\mathbf{x}_{t_i}^i - \mathbf{x}_{t_j}^j\|^2 \\ g(t_i, t_j - 1) + \|\mathbf{x}_{t_i}^i - \mathbf{x}_{t_j}^j\|^2 \end{cases}$$

(3) 終了: $d^2(X_i, X_j) = g(l_i, l_j) / (l_i + l_j)$

なお、 $\|\cdot\|$ はユークリッドノルムである。

3. カーネル PCA

文献 [16] に沿って、カーネル PCA を説明する。 $\mathcal{X} = \{X_1, X_2, \dots\}$ を有限あるいは無限集合 (必ずしもベクトル空間の部分集合に限らない) とし、 k を $\mathcal{X} \times \mathcal{X}$ 上で定義された正定値のカーネル関数とする。このとき、ヒルベルト空間 \mathcal{H} への写像 $\Phi: \mathcal{X} \rightarrow \mathcal{H}$ が存在して、任意の $X, X' \in \mathcal{X}$ に対して、 $k(X, X') = \langle \Phi(X), \Phi(X') \rangle$ が成立する。なお、 $\langle \cdot, \cdot \rangle$ はヒルベルト空間 \mathcal{H} の内積である。

カーネル PCA は集合 $\{\Phi(X_1), \Phi(X_2), \dots, \Phi(X_n)\}$ の主成分分析を行うものである。共分散行列

$C = \frac{1}{n} \sum_i \Phi(X_i) \Phi(X_i)^T$ の固有値を λ , 固有ベクトルを v とする: $Cv = \lambda v$. 0 でない固有値をもつ固有ベクトル v は $\{\Phi(X_i) \mid 1 \leq i \leq n\}$ が張る部分空間上にあり, $v = \sum_{i=1}^n \alpha_i \Phi(X_i)$ と展開できる. このとき, 展開係数 $\alpha = (\alpha_1, \dots, \alpha_n)^T$ は

$$n\lambda\alpha = K\alpha \quad (1)$$

すなわち, カーネル行列 $K: K_{ij} = k(X_i, X_j)$ の固有ベクトルとして求まる. 結局, $\Phi(X_i)$ の m 次主成分特徴は以下の式で計算できることが示せる.

$$\langle v^m, \Phi(X_i) \rangle = \sqrt{n\lambda_m} \alpha^m(i) \quad (1 \leq i \leq n) \quad (2)$$

ここに, λ_m は行列 K の固有値, $\alpha^m(i)$ は固有ベクトル α^m の i 番目の要素である. なお, 共分散行列 C の計算において, $\frac{1}{n} \sum_{i=1}^n \Phi(X_i) = \mathbf{0}$ を仮定したが, これはカーネル行列 K を中心化 (centering) することにより, 容易に実現できる.

4. ベクトル空間への埋込

4.1 手法 1: MDS

手法 1 では MDS [21] を用いて,

$$\|\Phi_1(X_i) - \Phi_1(X_j)\|^2 = d^2(X_i, X_j) \quad (1 \leq i, j \leq n) \quad (3)$$

を満たすような写像 $\Phi_1: \mathcal{X} \rightarrow \mathbb{R}^n$ を求める.

以下, $\Phi_1(X_i)$ を z_i と略記する. カーネル行列を中心化するために, $\bar{z} = \frac{1}{n} \sum_i z_i$ として, 内積 $k_1(X_i, X_j) = \langle z_i - \bar{z}, z_j - \bar{z} \rangle$ を考える.

$$\begin{aligned} & \langle z_i - \bar{z}, z_j - \bar{z} \rangle \\ &= \langle z_i, z_j \rangle - \frac{1}{n} \sum_{l=1}^n \langle z_l, z_j \rangle \\ & \quad - \frac{1}{n} \sum_{m=1}^n \langle z_i, z_m \rangle + \frac{1}{n^2} \sum_{l=1}^n \sum_{m=1}^n \langle z_l, z_m \rangle \end{aligned} \quad (4)$$

一般に, $z, z' \in \mathbb{R}^n$ に対して, $\|z - z'\|^2 = \langle z - z', z - z' \rangle = \langle z, z \rangle + \langle z', z' \rangle - 2\langle z, z' \rangle$ より, 以下が成立する.

$$\langle z, z' \rangle = -\frac{1}{2} \{ \|z - z'\|^2 - \|z\|^2 - \|z'\|^2 \} \quad (5)$$

式 (4) へ式 (5) を代入し, 整理すると, 以下が得られる.

$$\begin{aligned} & k_1(X_i, X_j) \\ &= -\frac{1}{2} \|z_i - z_j\|^2 + \frac{1}{2n} \sum_{l=1}^n \|z_i - z_l\|^2 \\ & \quad + \frac{1}{2n} \sum_{l=1}^n \|z_j - z_l\|^2 \\ & \quad - \frac{1}{2n^2} \sum_{l=1}^n \sum_{m=1}^n \|z_l - z_m\|^2 \end{aligned} \quad (6)$$

式 (6) に式 (3) を代入して, DTW 距離からカーネル行列を求める式が得られる.

$$\begin{aligned} & k_1(X_i, X_j) \\ &= -\frac{1}{2} d^2(X_i, X_j) + \frac{1}{2n} \sum_{l=1}^n d^2(X_i, X_l) \\ & \quad + \frac{1}{2n} \sum_{l=1}^n d^2(X_j, X_l) \\ & \quad - \frac{1}{2n^2} \sum_{l=1}^n \sum_{m=1}^n d^2(X_l, X_m) \end{aligned} \quad (7)$$

カーネル行列 $K: K_{ij} = k_1(X_i, X_j)$ の固有値解析を行い, 以下のように分解する.

$$K = U\Lambda U^T \quad (8)$$

ここで, 固有値行列 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \dots \geq \lambda_n$, 固有ベクトル行列 $U = [e^1, \dots, e^n]$ である.

行列 K が半正定値, すなわち $\lambda_n \geq 0$ である場合には, $Z = \Lambda^{\frac{1}{2}} U^T$ とおくと, $K = Z^T Z$ となるので, Z の i 列は, $z_i - \bar{z}$ であると考えられる. 重心位置に原点を移動したものを改めて z_i ($1 \leq i \leq n$) とする. 次元縮小のため, カーネル行列 K をフロベニウスノルムの意味で最もよく近似する (あるいは, $\{z_i \mid 1 \leq i \leq n\}$ に関する分散を最大限に保持する), n 次元空間から p 次元空間への射影を考える. その射影による z_i の像 \tilde{z}_i は, よく知られているように, カーネル行列 K の大きい方から p 個の固有値, 固有ベクトルを用いて, 以下のように表される.

$$\tilde{z}_i = (\sqrt{\lambda_1} e^1(i), \sqrt{\lambda_2} e^2(i), \dots, \sqrt{\lambda_p} e^p(i))^T \quad (1 \leq i \leq n) \quad (9)$$

なお, 上の式と式 (2) を比較すると, 定数倍 (\sqrt{n}) を除いて一致している.

さて、 \mathcal{X} がベクトル空間であり、 $d^2(\cdot, \cdot)$ がベクトル間のユークリッド距離である場合には、MDS は古典的スケーリングと呼ばれて、PCA と等価であり、行列 K が半正定値となることが知られている [23]。しかし、DTW 距離は、俗に距離とは呼ばれるものの、距離の公理の一つである三角不等式： $\forall x \forall y \forall z (d(x, z) \leq d(x, y) + d(y, z))$ を満たさないため、式 (7) から定まる行列 K は必ずしも半正定値とは限らない。手法 1 では、いささか強引ではあるが、負の固有値、固有ベクトルを無視して、式 (9) によりユークリッド空間への埋込を行う。

4.2 手法 2：擬似ユークリッド空間への埋込

前節と同様に、式 (7) を用いて DTW 距離から行列 K を求め、式 (8) のように固有値分解する。手法 1 では、行列 K をユークリッド空間における内積行列と考えることにより埋込を行ったが、手法 2 では、行列 K を擬似ユークリッド空間 $\mathfrak{R}^{(n^+, n^-)}$ における対称双一次形式の行列と考えることにより、負の固有値、固有ベクトルを除外することなく埋込を行う [7], [8], [13]。ここで、 n^+ 、 n^- は K の正、負の固有値の数を表している。

具体的な手続きは以下のとおりである。式 (8) の固有値行列 Λ の絶対値 $\bar{\Lambda}$ をとる： $\bar{\Lambda} = \text{diag}(|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|)$ 。 $Z = \bar{\Lambda}^{\frac{1}{2}} U^T$ とし、絶対値の大きい固有値から p 個選んで、埋込座標とする。

$$\tilde{z}_i = \left(\sqrt{|\lambda_{m_1}|} e^{m_1(i)}, \dots, \sqrt{|\lambda_{m_p}|} e^{m_p(i)} \right)^T \quad (1 \leq i \leq n) \quad (10)$$

擬似ユークリッド空間について補足説明する。擬似ユークリッド空間 $\mathfrak{R}^{(n^+, n^-)}$ は、ユークリッド空間の内積に対応する以下のような双一次形式： $\langle \cdot, \cdot \rangle_M$ をもつベクトル空間である。

$$\langle z, z' \rangle_M = z^T M z' \quad (11)$$

$$M = \begin{pmatrix} I_{n^+} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -I_{n^-} \end{pmatrix}_{n \times n}$$

M は擬似ユークリッド空間の符号定数行列と呼ばれる。擬似ユークリッド空間には、双一次形式 $\langle \cdot, \cdot \rangle_M$ から導かれる擬似距離： $\|z - z'\|_M^2 = (z - z')^T M (z - z')$ が定義される。手法 2 では、

$$\|\Phi_2(X_i) - \Phi_2(X_j)\|_M^2 = d^2(X_i, X_j)$$

$$(1 \leq i, j \leq n) \quad (12)$$

を満たすような写像 $\Phi_2 : \mathcal{X} \rightarrow \mathfrak{R}^{(n^+, n^-)}$ を求めるのである。

なお、パターン認識の世界では、特徴空間とは一般にヒルベルト空間（有限次元の場合には、ユークリッド空間とも呼ぶ）を意味するが、手法 2 は擬似ユークリッド空間をも特徴空間と考える立場である [7]。

4.3 手法 3：ラプラシアン固有マップ法

手法 3 では、ラプラシアン固有マップ法 [2], [3] を用いて、時系列の空間埋込と次元縮小を行う。以下、ラプラシアン固有マップ法のアルゴリズムを示す。

(1) 類似度行列 W を計算する。

$$W_{ij} = \begin{cases} e^{-d^2(X_i, X_j)/t} & i \neq j \wedge X_j \in Nbr(X_i) \\ 0 & \text{上記以外} \end{cases} \quad (13)$$

ここに、 $X_j \in Nbr(X_i)$ は X_j が X_i の近傍にあるという意味であり、 $t (> 0)$ はハイパパラメータである。

(2) ラプラシアン行列 L を計算する。

$L = D - W$ 、 D は、 $D_{ii} = \sum_{j=1}^n W_{ij}$ なる対角行列である。なお、5.2 で示すように、行列 L は半正定値である。

(3) 一般化固有値問題を解く。

一般化固有値問題 $Le = \lambda De$ に対して、固有値の小さなものから固有ベクトル e^1, \dots, e^n ($\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$) を計算する。

(4) 埋込座標を計算する。

$U = [e^1, e^2, \dots, e^n]$ の i 行を、埋込座標ベクトル z_i とする。すなわち、 $Z = [z_1, \dots, z_n] = U^T$ 。固有値の小さなものから p 個の固有ベクトルを選んで、

$$\tilde{z}_i = (e^1(i), e^2(i), \dots, e^p(i))^T \quad (1 \leq i \leq n) \quad (14)$$

なお、ラプラシアン行列 L の擬似逆行列 $K_L = L^\dagger$ がラプラシアン固有マップ法に対応するカーネル行列とみなせるようである [9]。したがって、 L の小さい固有値をもつ固有ベクトルを用いての埋込は、 K_L をカーネル行列とするカーネル PCA と等しいといえる。

5. 埋込前・後のデータ間距離

ここでは、時系列データ間距離がベクトル空間へ埋込込むことにより、どのように変化しているかを考察する。

5.1 手法 1, 2 の場合

手法 1, 2 では, 式 (3), 式 (12) から明らかなように, 短い DTW 距離も長い DTW 距離も同様に保存する大域的埋込を行っているといえる.

5.2 手法 3 の場合

一般化固有値問題 $Le = \lambda De$ の解の意味について考察する.

まず, $\sum_i \sum_j |z_i - z_j|^2 W_{ij}$ を最小とするような, 時系列データから一次元空間への写像 $\Psi: \mathcal{X} \rightarrow \mathcal{R} (X_i \mapsto z_i)$, すなわち, 近いデータ同士はできる限りそのまま近くへ写されるような写像を求める. $u = (z_1, \dots, z_n)^T$ とすると, 以下が成り立つ.

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n |z_i - z_j|^2 W_{ij} = u^T L u \quad (15)$$

式 (15) は $\sum_i \sum_j |z_i - z_j|^2 W_{ij}$ を以下のように表すことにより導かれる.

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n (z_i^2 + z_j^2 - 2z_i z_j) W_{ij} \\ &= \sum_{i=1}^n z_i^2 D_{ii} + \sum_{j=1}^n z_j^2 D_{jj} - 2 \sum_{i=1}^n \sum_{j=1}^n z_i z_j W_{ij} \\ &= 2u^T L u \end{aligned} \quad (16)$$

したがって, $u^T D u = 1$ を u に関する拘束条件として, 以下のような条件付最小化問題を考えればよい.

$$u^* = \operatorname{argmin}_{u^T D u = 1} u^T L u \quad (17)$$

式 (15) より, L は半正定値行列で, 目的関数を最小化するベクトル u^* は一般化固有値問題 $Le = \lambda De$ の最小固有値に対応する固有ベクトルであることが分かる.

ここまでは一次元空間への埋込を考えたが, これを p 次元空間へ拡張すると, $n \times p$ 行列 $\tilde{U} = [u^1 u^2 \dots u^p]$ の i 行が \tilde{z}_i となる. すなわち, $\tilde{z}_i = (u^1(i), u^2(i), \dots, u^p(i))^T$. とすることで,

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\tilde{z}_i - \tilde{z}_j\|^2 W_{ij} = \operatorname{tr}(\tilde{U}^T L \tilde{U}) \quad (18)$$

なので, 以下の最小化問題を解けばよい.

$$\tilde{U}^* = \operatorname{argmin}_{\tilde{U}^T D \tilde{U} = I} \operatorname{tr}(\tilde{U}^T L \tilde{U}) \quad (19)$$

この最小化問題は, よく知られているように (例えば,

[19]), 一般化固有値問題 $Le = \lambda De$ の小さい方から p 個の固有値と固有ベクトルを求める問題に帰着する. 以上より, 手法 3 におけるラプラシアン固有マップ法での一般化固有値問題は, 式 (18) の最小化問題と等価であるということが分かる.

さて, 式 (18) において, 埋め込まれた後の距離が類似度 W_{ij} によって重み付けがなされているため, 手法 3 は短い距離ほど重点的に保存される局所的埋込であるといえる.

6. 分類

6.1 分類器の学習

分類器としては様々なものが考えられるが, ここでは簡単な線形識別関数を考える.

$$f(X) = \langle w, \tilde{\Phi}(X) \rangle + b = \langle w, \tilde{z} \rangle + b \quad (20)$$

なお, 手法 2 については, $\langle w, \tilde{\Phi}(X) \rangle_M$ とすべきであるが, $w' = M w$ として, $\langle w, \tilde{z} \rangle_M = \langle w', \tilde{z} \rangle$ を考えることにより, これ以降はユークリッド空間の場合と同様に扱う.

識別関数の学習において, 本来は, 擬似ユークリッド空間の位相的構造を考慮すべきであろうが, それは困難である^(注1). したがって, ここでは簡単のため, (1) 特徴ベクトルの抽出は擬似ユークリッド空間にて行い, (2) 抽出された特徴ベクトルによる分類はユークリッド空間にて行うこととした. 手法 1 と対比させれば, これは固有値の絶対値が大きな固有ベクトルを用いた埋込及び分類ともみなせる. なお, 擬似ユークリッド空間を用いた関連研究 [8], [13] でも同様な簡略化が行われている.

6.1.1 最小二乗誤差

ラベル付データ $\{(\tilde{\Phi}(X_i), y_i) \mid 1 \leq i \leq s\}$ に関する二乗誤差を最小とする識別境界を決定する. つまり, 以下の誤差関数,

$$\operatorname{Err}(w, b) = \sum_{i=1}^s \{y_i - (\langle w, \tilde{z}_i \rangle + b)\}^2 \quad (21)$$

の w, b に関する最小化を考えればよい.

この解はよく知られているように (例えば, [11]), $\hat{z}_i^T = (1, \tilde{z}_i^T) (1 \leq i \leq s)$, $\hat{w}^T = (b, w^T)$ とおけば,

(注1): 擬似ノルムは位相空間を導かない [7] ため, 関数の連続性の概念が導入できない, あるいは, より具体的に, マージン最大化の場合の $\|w\|$ の値が負になるなどのため.

$$(\hat{w})^* = (\hat{Z}_{lab})^\dagger y \quad (22)$$

ここに、 $\hat{Z}_{lab} = [\hat{z}_1 \dots \hat{z}_s]$ 、 $y^T = (y_1, \dots, y_s)$ である。

6.1.2 最大マージン

SVM,あるいはOHC(Optimum Hyperplane classifier)と呼ばれる分類器ではマージンを最大化する超平面により分類を行う[5]。マージンを最大にする超平面は、以下の条件のもと、 $\|w\|^2$ を最小化することにより得られる。

$$y_i(\langle w, \tilde{z}_i \rangle + b) \geq 1, \quad i = 1, \dots, s \quad (23)$$

これは、以下の双対問題(二次計画法)の、 $\alpha = [\alpha_1, \dots, \alpha_s]^T$ に関する最小化を考えるのと同様である。

$$\begin{aligned} \min_{\alpha} W(\alpha) \\ = \min_{\alpha} \left\{ \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T Y^T \tilde{Z}_{lab}^T \tilde{Z}_{lab} Y \alpha \right\} \\ \text{subject to } \alpha_i \geq 0 (\forall i) \quad \mathbf{1}^T Y \alpha = 0 \end{aligned}$$

ここで、 $\tilde{Z}_{lab} = [\tilde{z}_1 \dots \tilde{z}_s]$ 、 $Y = \text{diag}(y)$ であり、 $\mathbf{1}$ はすべての要素が1からなる s 次元ベクトルである。

さて、 $W(\alpha)$ 中の $\tilde{Z}_{lab}^T \tilde{Z}_{lab}$ は、 α に関する二次形式のヘッセ行列とみなせるが、この行列は $K = U \Lambda U^T$ (手法1)、 $\bar{K} = U \bar{\Lambda} U^T$ (手法2)の、ラベル付データに対応する要素からなる部分行列である。

6.2 ラベルなしデータの分類

$f(X) = \langle w^*, \tilde{z} \rangle + b^*$ とし、 $X_i (i > s)$ について、

$$y_i = \begin{cases} 1, & \text{if } f(X_i) \geq 0 \\ -1, & \text{if } f(X_i) < 0 \end{cases} \quad (24)$$

7. 実験

提案アプローチの有効性を検証するために、実験を行う。実験1では三つの埋込手法の比較を行い、実験2では k -近傍法との比較を行う。

7.1 実験1

オーストラリア手話データ[10]を用いて3手法の比較を行う。オーストラリア手話データは、5人の被験者から得た95の手話単語データであり、それぞれ、9次元特徴ベクトル(注2)から構成される、平均51フレームの長さの時系列データである。

実験では95単語のなかから、2組のペア「sad」と「what」、「go」と「please」を取り出し、2クラス分類

を行った。いずれも右手だけを用いる片手手話単語であるが、「sad」と「what」は手の形が異なるのに対し、「go」と「please」は手の形が同じである(注3)。データ数は各単語それぞれ70ずつである。

なお、手法3において類似度行列を定義する際、第8近傍までを近傍に含まれるものとした。また、 t の値は、実験的に決め、ここでは両方とも $t = 10000$ を用いた。

ところで、手法1と手法2では、固有値の正負の分布が結果に影響してくると思われる。そこで、負の固有値の数を調べてみると、「sad」と「what」、「go」と「please」ともにそれぞれ140個の固有値中80個前後が負の値であった。負の固有値が多く存在していることから、DTW距離は「たちの悪い」距離であることが分かる。

7.1.1 分類実験

ラベル付データ数を10, 20, ..., 130と、10ずつ変化させ、ランダムに選択した。それぞれ30回分類実験を行い、その平均分類誤り率を見た。埋込次元数 p は、ラベル付データ数の20%とした。なお、分類器として、ここでは関連研究[3],[8],[13]で良い結果が報告されている、マージンを最大化する(手法1, 2)、及び二乗誤差を最小化する(手法3)線形識別器を用いた。

図1と図2に、ラベル付データ数を変化させたときの平均分類誤り率を、それぞれ示す。なお、図中の破線(Euclid)は手法1、点線(p-Euclid)は手法2、

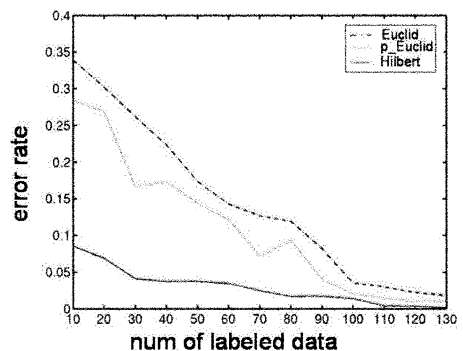


図1 実験1:「sad」vs.「what」の結果
Fig.1 Results of Ex.1 with “sad” and “what.”

(注2): 被験者の右手に装着した任天堂パワーグローブから、右手の3次元座標、ロール(手の平の向き)、5本の指の角度を得ている。

(注3): 一般に手話単語の表現は、(1)右手と左手の使い方(片手、両手対称、両手独立など)、(2)手の形(平手、げんこつなど指の曲げ方)、(3)手の動かし方により行う。

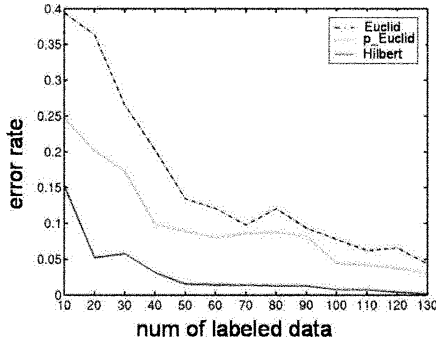


図 2 実験 1: 「go」 vs. 「please」の結果
Fig. 2 Results of Ex.1 with “go” and “please.”

実線 (Hilbert) は手法 3 である。いずれの分類問題についても、手法 1 よりも手法 2 の方が精度が高いことが図より確認できる。また、手法 3 の精度が最も高いことも確認できる。以下にその理由を考察する。

もともと、DTW 距離はパターンマッチング問題を考えるためのものであるため、短い距離は信頼性があるが、長い距離は意味をもたないといった性質があると考えられる。また、5. で説明したように、手法 1, 2 は短い距離も長い距離も平等に埋め込む手法であるのに対し、手法 3 は短い距離に重点をおいて埋込を行っている。したがって、上記の DTW 距離の性質を反映しているといえる手法 3 が手法 1, 2 より精度が高かったのではないかと考えられる。

各手法の埋込前後のデータ間距離に関する分析を確認するため、「go」と「please」の計 140 データを対象に、各データ間の埋込前の DTW 距離 (X 軸) と埋込後の距離 (Y 軸) をプロットしてみた。図 3 が手法 1 (手法 2 もよく似た図になる)、図 4 が手法 3 である。図 3 は直線状のグラフであり、DTW 距離と埋込後の距離がほぼ比例し、短い DTW 距離も長い DTW 距離も同様に保存されていることが分かる。図 4 は扇状に広がったグラフであり、DTW 距離が長くなるにつれて、同じ DTW 距離に対応する埋込距離の差異が拡大しており、長い DTW 距離が保存されていないことが分かる。

7.2 実験 2

次に、三つの手法のうち最も精度の高かった手法 3 を、同じ距離ベースの手法であり、同じ DTW 距離を使うものの、埋込を行わない k -近傍法 ($k = 1, 3, 5$) と比較する。実験条件は実験 1 と同様である。

まず、実験 1 と同じ手話データを用いた場合の、ラ

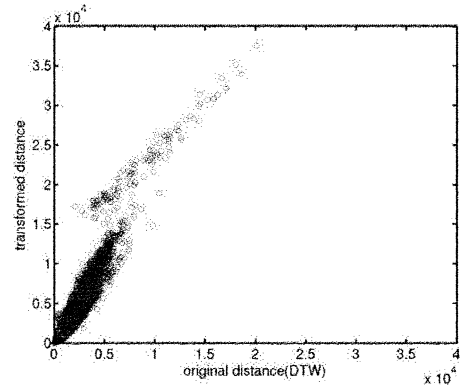


図 3 手法 1: DTW 距離と埋め込み後の距離
Fig. 3 Method 1: Original and transformed distances.

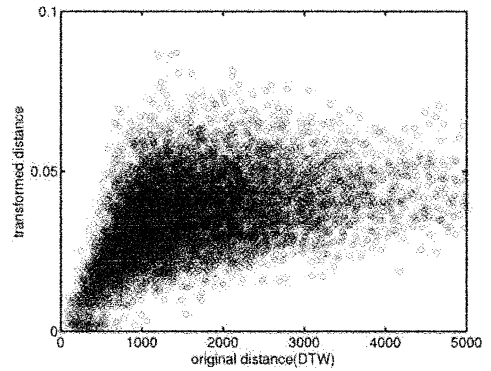


図 4 手法 3: DTW 距離と埋め込み後の距離
Fig. 4 Method 3: Original and transformed distances.

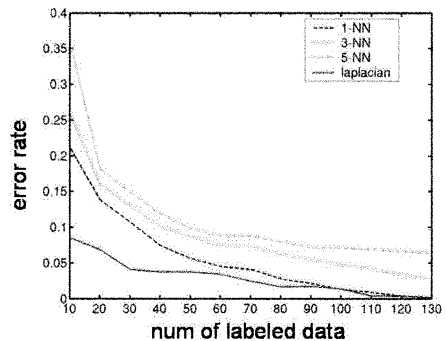


図 5 「sad」 vs. 「what」の結果
Fig. 5 Results of Ex.2 with “sad” and “what.”

ベル付データ数を変化させたときの平均分類誤り率を図 5 と図 6 に示す。なお、図中の実線 (laplacian) は手法 3、破点線 (1-NN) は 1-近傍法、点線 (3-NN) は 3-近傍法、破線 (5-NN) は 5-近傍法である。

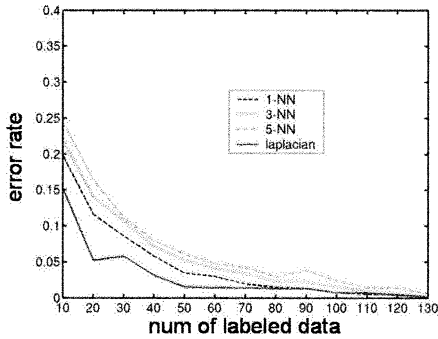


図 6 実験 2: 「go」 vs. 「please」の結果
Fig. 6 Results of Ex.2 with “go” and “please.”

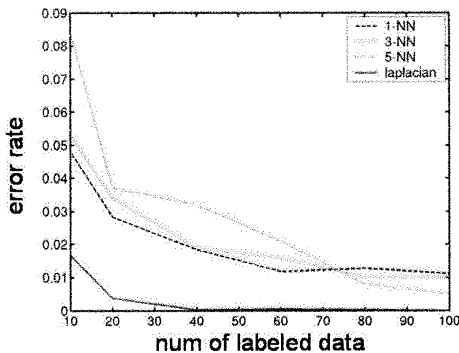


図 7 実験 2: Control Chart データの結果
Fig. 7 Results of Ex.2 with control chart.

また、Control Chart Time Series [10] を用いて、同様な実験を行った。データベースは 6 種類の異なった波形パターンからなり、各種類ごとに 100 個のスカラの時系列データがある。その中から最も類似している 2 種類の波形、Upward shift (上昇しているが突然シフトするデータ) と Increasing trend (常に上昇傾向であるデータ) を選んで、 200×200 の DTW 距離行列を作成し、2 クラス分類を行った。

図 7 にラベル付データ数を変化させたときの k -近傍法 ($k = 1, 3, 5$) との平均分類誤り率の比較結果を示す。なお、図中の実線 (laplacian) は手法 3、破点線 (1-NN) は最近傍法、点線 (3-NN) は 3-近傍法、破線 (5-NN) は 5-近傍法である。なお、 $t = 10000$ を用いた。

手話データ (図 5 と図 6)、Control Chart Time Series (図 7) の分類結果より、手法 3 の精度が最も高く、 k -近傍法よりも優れていることが確認できる。以下に、その理由を考察する。手法 3 では、線形識別器を用いて分類を行ったため、ノイズに頑健であったと

考えられる。更に、手法 3 では、 k -近傍以外のデータ (ラベルのあるなしにかかわらず、8 近傍まで) との DTW 距離も考慮された上で、埋込座標が求まっているため、ラベルなしデータが有効利用されている [17] と考えられる。

8. む す び

DTW (動的時間伸縮) により得られたデータ間距離をもとに、時系列データをベクトル空間へ埋め込み、埋込空間にて分類するアプローチを提案した。ラベル付データとラベルなしデータが最初に与えられる学習問題設定のもとで、埋込手法の候補として、3 手法を考えた。DTW 距離の性質と埋込手法の適合性に関する考察、及び分類実験により、ラプラシアン固有マップ法による埋込が 3 手法の中で最も高い分類精度につながる事が分かった。また、ラプラシアン固有マップ法を用いた提案アプローチは k -近傍法より高い分類精度を実験で示した。

今後の研究課題として、新規時系列の埋込 [4]、フィッシャーカーネル法 [12] との比較、クラスタリング問題への応用などが挙げられる。

文 献

- [1] C. Bahlmann, B. Haasdonk, and H. Burkhardt, “Online handwriting recognition with support vector machines—a kernel approach,” Proc. 8th Int. W/Shop on Frontiers in Handwriting Recognition, pp.49–54, 2002.
- [2] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” Advances in Neural Information Processing Systems, vol.14, pp.585–591, 2002.
- [3] M. Belkin and P. Niyogi, “Using manifold structure for partially labelled classification,” Advances in Neural Information Processing Systems, vol.15, pp.929–936, 2003.
- [4] Y. Bengio, P. Vincent, and J.F. Paiement, “Spectral clustering and kernel PCA are learning eigenfunctions,” Technical Report 1239, Departement d’informatique et recherche operationnelle, Universite de Montreal, 2003.
- [5] C. Corres and V. Vapnik, “Support vector networks,” Mach. Learn., vol.20, pp.273–297, 1995.
- [6] T. Darrell and A. Pentland, “Space-time gestures,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.335–340, 1993.
- [7] L. Goldfarb, “A new approach to pattern recognition,” in Progress in Pattern Recognition, vol.2, pp.241–402, Elsevier Science Publishers B.V., 1985.
- [8] T. Graepel, R. Herbrich, P. Bollmann-Sdorra, and

- K. Obermayer, "Classification on pairwise proximity data," *Advances in Neural Information Processing*, vol.11, pp.438-444, 1999.
- [9] J. Ham, D.D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," TR-110, Max-Planck-Institut für biologische Kybernetik, Tübingen, 2003.
- [10] S. Hettich and S.D. Bay, UCI Repository of KDD Databases, <http://kdd.ics.uci.edu/> 1999.
- [11] 石井健一郎, 上田修功, 前田英作, 村瀬 洋, わかりやすいパターン認識, オーム社, 1998.
- [12] T. Jaakkola, M. Diekhans, and D. Haussler, "Using the Fisher kernel method to detect remote protein homologies," *Proc. 7th Int. Conf. on Intelligent Systems for Molecular Biology*, pp.149-158, 1999.
- [13] E. Pekalska, P. Paclik, and R.P.W. Duin, "A generalized kernel approach to dissimilarity-based classification," *J. Machine Learning Research, Special Issue on Kernel Methods*, vol.2, no.2, pp.175-211, 2002.
- [14] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [15] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol.290, pp.2323-2326, 2000.
- [16] B. Schölkopf, A.J. Smola, and K.R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol.10, pp.1299-1319, 1998.
- [17] M. Seeger, "Learning with labeled and unlabeled data," Technical Report, Institute for Adaptive and Neural Computation, University of Edinburgh, 2001.
- [18] H. Shimodaira, K. Noma, M. Nakai, and S. Sagayama, "Dynamic time-alignment kernel in support vector machine," *Neural Information Processing Systems*, vol.14, pp.921-928, 2002.
- [19] G. Strang, *Linear Algebra and Its Applications*, Harcourt Brace Javanovich College Publishers, 1988.
- [20] J. Tenenbaum, V. de Silba, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol.290, pp.2319-2323, 2000.
- [21] W.S. Torgerson, *Theory and methods of scaling*, John Wiley & Sons, 1958.
- [22] P. Vincent and Y. Bengio, "K-local hyperplane and convex distance nearest neighbor algorithms," *Advances in Neural Information Processing Systems*, vol.14, pp.985-992, 2002.
- [23] C. Williams, "On a connection between kernel PCA and metric multidimensional scaling," *Advances in Neural Information Processing Systems*, vol.13, pp.675-681, 2001.
- [24] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov models," *Proc. Computer Vision and Pattern Recognition*, pp.379-385, 1992.
- [25] 大和淳司, 上田修功, 和田俊和, "動作認識のための状態

遷移モデル," *人工知能誌*, vol.17, no.1, pp.41-46, 2001.
(平成 16 年 4 月 26 日受付, 8 月 9 日再受付)



水原 悠子

2002 広島市大・情報科学卒. 2004 同大学院情報科学研究科修士課程了. 2004 パナソニック AVC マルチメディアソフト(株)入社.



林 朗 (正員)

1974 京大・理卒. 同年, 日本アイビーエム(株)入社. 1988 ブラウン大学計算機科学科修士課程了. 1991 テキサス大学オースチン校計算機科学科博士課程了(PhD.). 九州工業大学情報工学部客員助教授を経て, 現在は広島市立大学情報科学部教授. 情報処理学会, 人工知能学会, 日本ロボット学会, AAAI, ACM, IEEE 各会員.



末松 伸朗

1988 九大・理・物理卒. 1990 同大学院修士課程了. 同年, (株)富士通研究所入社. 1994 広島市立大学情報科学部助手, 2002 同助教授, 博士(工学). 人工知能学会, 情報処理学会, 日本認知科学会各会員.