H-003

半正定値計画法による時系列データの埋め込み

Embedding Time Series Data by Solving the Semidefinite Programming Problem

成田 浩之 † 林 朗 † 末松 伸朗 † 岩田 一貴 †
Hiroyuki Narita Akira Hayashi Nobuo Suematsu Kazunori Iwata

1 はじめに

動画，楽曲，音声，センサーデータなどの時系列データ解析の必要性が高まっている．時系列データの多くは長さの異なる非ベクトルデータであり，そのマッチングには三角不等式を満たさない擬距離である動的時間伸縮 (DTW) 距離がしばしば用いられてきた．時間軸の固定されたユークリッド距離に対し，DTW 距離は時間軸のズレに頑健であり，より直感的な類似性を反映していると考えられる．

DTW 擬距離を距離とみなし，時系列データを距離を保存するようにユークリッド空間へ埋め込み，ベクトルデータに変換する．ベクトルデータに変換することにより，可視化ばかりでなく，カーネル法などのベクトルデータを対象としたパターン認識手法の適用が可能になり，分類精度の向上が期待できるからである．

距離が与えられたデータ集合のユークリッド空間への埋め込みには，多次元尺度法 (MDS)[1] がよく知られている．しかし．DTW 距離は擬距離であるため．MDS による埋め込みではカーネル行列の半正定値性を満たせず，結果的に大きな誤差が発生する [2].

本研究では，上記問題点を持たない埋め込み手法として，半正定値計画法による埋め込みを提案する．DTW 距離から求めるカーネル行列の半正定値性を拘束条件として，埋め込み前と埋め込み後の距離の差を最小化する最適化問題を解く．

実データを用いた埋め込み実験，および，埋め込み空間における分類実験により，提案手法は従来手法に比べて，より精度の高い埋め込み，および分類が可能であることがわかった．

2 DTW(動的時間伸縮)

DTW（Dynamic Time Warping）は 2 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全ての対応を評価し，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 [3]．本論文で用いた時系列 X_i, X_j 間 DTW 距離 $d^2(i,j)$ は以下のように求めた．ただし，$X_i = (\boldsymbol{x}_1^i, \ldots, \boldsymbol{x}_{l_i}^i)$, $X_j = (\boldsymbol{x}_1^j, \ldots, \boldsymbol{x}_{l_j}^j)$ とする．

† 広島市立大学 大学院情報科学研究科
〒731-3194 広島市安佐南区大塚東 3-4-1
Graduate School of Infomation Sciences, Hiroshima City University, 3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima-shi,731-3194 Japan
Email: narita@robotics.im.hiroshima-cu.ac.jp

1. 拘束条件
   - 始点：$g(1,1) = 0$
   - 終点：$g(l_i, l_j)$
   - 境界条件：$g(t_i, 0) = g(0, t_j) = \infty$
2. 再帰的計算

   $1 \le t_i \le l_i$, $1 \le t_j \le l_j$ において

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

3. 終了： $d^2(i,j) = g(l_i, l_j)$

なお，$\|\cdot\|$ はユークリッドノルムである．

3 半正定値計画法

半正定値計画法（Semidefinite programming : SDP）とは，以下の目的関数を最小にするような半正定値行列 $\boldsymbol{X}$ を最適化法によって求める解く手法である [4].

[SDP]

$$\begin{array}{c} \min_{\boldsymbol{X}} \ \boldsymbol{C} \bullet \boldsymbol{X} \\ s.t.\ \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i \quad (i = 1, \ldots, m) \\ \boldsymbol{X} \in \boldsymbol{S}_n^+, \boldsymbol{C} \in \boldsymbol{S}_n, \\ \boldsymbol{A}_i \in \boldsymbol{S}_n, b_i \in \boldsymbol{R} \end{array} \quad (2)$$

ここに，$\boldsymbol{S}_n, \boldsymbol{S}_n^+$ はそれぞれ $n \times n$ 実対称行列，半正定値対称行列の集合である．また，$\bullet$ は行列間の内積，$A \bullet B = \Sigma_{ij} A(i,j) \cdot B(i,j)$ を表す．m は拘束条件の数である．

なお，本研究では SDP の拡張である一般化凸錐計画（General Convex Cone Programming: GCCP）を用いる．

[GCCP]

$$\begin{array}{c} \min_{\boldsymbol{X}_j, \boldsymbol{x}_i, z} \ \sum_{j=1}^{n_s} \boldsymbol{C}_j \bullet \boldsymbol{X}_j + \sum_{j=1}^{n_q} \boldsymbol{c}_j \cdot \boldsymbol{x}_j + gz \\ s.t. \ \sum_{j=1}^{n_s} \boldsymbol{A}_{rj} \bullet \boldsymbol{X}_j + \sum_{j=1}^{n_q} \boldsymbol{a}_{ri} \cdot \boldsymbol{x}_i + g_r z = b_r, \ \forall_r \\ \boldsymbol{X}_j \in \boldsymbol{S}_n^+, \boldsymbol{x}_i \in \boldsymbol{Q}^{p_i}, z \in \boldsymbol{P}^p, \\ \boldsymbol{A}_{rj}, \boldsymbol{C}_j \in \boldsymbol{S}_n, c_i, a_{rj} \boldsymbol{R}^q, \\ g, g_r \in \boldsymbol{R}^p, b_r \in \boldsymbol{R}^l \end{array} \quad (3)$$

ここに

$\boldsymbol{R}^p$：p 次元ユークリッド空間

$\boldsymbol{P}^p$：$\boldsymbol{R}^p$ におけるすべての座標が非負の象限．すなわち，すべての要素が非負の p 次元ベクトル集合

$\boldsymbol{Q}^q$：q 次元の 2 次錐．すなわち，$\boldsymbol{x}(1) \ge [\sum_{i=2}^{q} \boldsymbol{x}(i)^2]^{1/2}$ を満たす $\boldsymbol{x} = (\boldsymbol{x}(1), \ldots, \boldsymbol{x}(q)) \in \boldsymbol{R}^q$ のベクトル

4 提案手法

4.1 時系列データの埋め込み

1. 与えられた n 個の時系列データ $X_1, \ldots, X_n$ 間の DTW 距離 $\{d_{ij} | 1 \leq i, j \leq n\}$ を計算する.
2. DTW 距離はパターンマッチングスコアであり，短い距離（近傍内距離）は信頼できるのに対し，長い距離は信頼できない．短い距離（近傍関係）に注目して埋め込みを行う方がよいことがすでにわかっている [2]. そこで，DTW 距離に基づいて，時系列データ間に近傍関係 ($\sim$) を導入する．たとえば，$d_{ij} < \epsilon$ ならば $X_i \sim X_j$ (ϵ 近傍)，あるいは X_j が X_i の k 近傍に含まれる（あるいは，その逆）ならば，$X_i \sim X_j$ (k 近傍) とする．最後に，近傍関係をまとめて，$\Omega = \{(i,j) | X_i \sim X_j, 1 \leq i, j \leq n\}$ とする.
3. 半正定値計画法を用いて近傍間 DTW 距離を保持し，半正定値性を満たすようなカーネル行列 $\boldsymbol{K} \succeq 0$ を求める [5]. ここで，$\boldsymbol{K} \succeq 0$ は $\boldsymbol{K} \in \boldsymbol{S}_n^+$ の略記である．目的関数は以下のようになる

$$\min_{\boldsymbol{K} \succeq 0} \sum_{(i,j) \in \Omega} \boldsymbol{w}_{ij} | d_{ij}^2 - \boldsymbol{B}_{ij} \bullet \boldsymbol{K} | + \lambda \, tr(\boldsymbol{K}) \quad (4)$$

$\boldsymbol{w}_{ij}$ は重み，$\boldsymbol{B}_{ij}$ はカーネル $\boldsymbol{K}$ から距離の二乗を求めるための行列，すなわち，$B_{ij}(i,i) = B_{ij}(j,j) = 1, B_{ij}(i,j) = B_{ij}(j,i) = -1$, その他の要素はすべて 0 の行列である．目的関数の第 2 項は正規化項であり，$tr(\boldsymbol{K})$ は埋め込み空間におけるデータの分散を示し，λ は正規化項の重みパラメータである．(4) は以下の GCCP に変換できる.

$$\begin{aligned} &\min_{\boldsymbol{K} \succeq 0, \boldsymbol{u}^+, \boldsymbol{u}^- \geq 0} \boldsymbol{w} \cdot (\boldsymbol{u}^+ + \boldsymbol{u}^-) + \lambda \boldsymbol{I}_n \bullet \boldsymbol{K} \\ &s.t. \quad d_{i,j}^2 - \boldsymbol{B}_{i,j} \bullet \boldsymbol{K} + \boldsymbol{u}^+ - \boldsymbol{u}^- = 0, \forall_r, \\ &\quad \boldsymbol{K} \in \boldsymbol{S}_n^+, \boldsymbol{u}^+ \in \boldsymbol{P}_m^+, \boldsymbol{u}^- \in \boldsymbol{P}_m^- \end{aligned} \quad (5)$$

ここで，m は ϵ 近傍の要素数，$\boldsymbol{P}_m^+$，$\boldsymbol{P}_m^-$ は m 次元ユークリッド空間における非負象限，負象限である.
4. 半正定値計画法によって得られたカーネル行列 $\boldsymbol{K}$ を固有値解析する．時系列 X_l の p 次元埋め込み座標ベクトル $\boldsymbol{z}(l)$ は $\boldsymbol{K}$ の固有値 λ_k，固有ベクトル $\boldsymbol{e}_k$ を降順に p 個用いて，式 (6) のようになる.

$$\boldsymbol{z}_k(l) = \sqrt{\lambda_k} \boldsymbol{e}_k^l \quad (1 \leq k \leq p), (1 \leq l \leq n) \quad (6)$$

ここで e_k^l は固有ベクトル $\boldsymbol{e}_k$ の l 番目の要素である.

4.2 新規データの埋め込み

新規データが与えられたときには，再度半正定値計画問題を解き，カーネル行列を計算すれば良いのであるが，それは計算負荷が大きい．ここでは，既存のデータに対してすでに求まっているカーネル行列を拡大することにより，新規データを含めたカーネル行列を求める方法を示す.

4.1 で求めたカーネル行列 $\boldsymbol{K}_n$ を用いて，拡大カーネル行列 $\tilde{\boldsymbol{K}}_{n+1}$ を以下の式 (7) と定義する．ここで，$\boldsymbol{b} \in \boldsymbol{R}^n, c \in \boldsymbol{R}$ である.

$$\tilde{\boldsymbol{K}}_{n+1} = \begin{bmatrix} \boldsymbol{K}_n & \boldsymbol{b}^T \\ \boldsymbol{b} & c \end{bmatrix} \succeq 0 \quad (7)$$

そして，以下の最適化問題を求める事で，新規時系列データを同じ空間に埋め込む事ができる.

$$\begin{aligned} &\min_{c \geq 0, \boldsymbol{b}} w_i | d_{i,n+1} - \boldsymbol{B}_{i,n+1} \bullet \boldsymbol{K}_{n+1} | \\ &s.t. \quad \boldsymbol{b} \in Range(\boldsymbol{K}_n), c - \boldsymbol{b}^T \boldsymbol{K}_n^+ \boldsymbol{b} \geq 0 \end{aligned} \quad (8)$$

4.3 分類

1. 分類器の構築

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用いる．ラベル付きデータ $\{(z(i), y_i) | y_i \in \pm 1, 1 \leq i \leq s\}$ を用いて，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,

$$\mathrm{Err}(\boldsymbol{w}, b) = \sum_{i=1}^{s} \{ y_i - (\langle \boldsymbol{w}, \boldsymbol{z}(i) \rangle + b) \}^2 \quad (9)$$

の $\boldsymbol{w}, b$ に関する最小化を考えればよい.

2. ラベルなしデータの分類

$\tilde{\boldsymbol{z}}_i (i > s)$ が，ラベルなしデータの場合,

$$y_i = \begin{cases} 1, & \text{if } \ y_i - \langle \boldsymbol{w}^*, \boldsymbol{z}(i) \rangle + b^* \geq 0 \\ -1, & \text{if } \ y_i - \langle \boldsymbol{w}^*, \boldsymbol{z}(i) \rangle + b^* < 0 \end{cases} \quad (10)$$

以上のアルゴリズムにより，時系列データを空間へ埋め込み，その埋め込んだ空間にて分類が行える

5 実験

提案手法の有効性を検証するために，実験を行う．実験 1 では埋め込みの精度を検証する実験を行い，実験 2，3 では分類精度を検討する.

実験 2，3 の違いを説明する．実験 2 では，半教師つき学習を行う．すなわち，ラベルつき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下で，ラベルつきデータ，ラベルなしデータの両方を最初に埋め込み，識別関数の学習とラベルなしデータの分類を行う．実験 3 では，教師付き学習を行う．すなわち，ラベルつきデータを最初に埋め込み，識別関数を学習し，その後，ラベルなしデータを新規データとして埋め込み，分類を行う.

[実験データ]

オーストラリア手話データ [8] 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，九次元特徴ベクトルから構成される，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言語であるが，「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．データ数は各単語それぞれ 70 ずつである.

5.1 実験 1

「go」と「please」のデータを使い，MDS，ISOMAP[6]，提案手法を用いて埋め込み実験を行った．ISOMAP は近傍関係に基づき計算された測地線距離に基づきカーネル行列を計算する手法であり，MDS の拡張と見なせる．提案手法のパラメータは $\lambda = -0.1$, $\boldsymbol{w}_{ij} = 1$，保持する近傍関係 $\epsilon = 31$ とした.

まず，MDS，ISOMAP，提案手法の固有値解析について触れる．各種法によって得られたカーネル行列を固有値解析したものを図 1 に示す．図 1 より，MDS，ISOMAP の右半分は負の固有値である事がわかる．対照的に，提案手法では負の固有値が出ずに，固有値が 0 に近づいていくことがわかる．そして，埋め込み後のカーネル行列から求めた距離を $\hat{d}_{ij}(K)$ とし，その誤り率を式 (11) で求め，表 1 に示す.

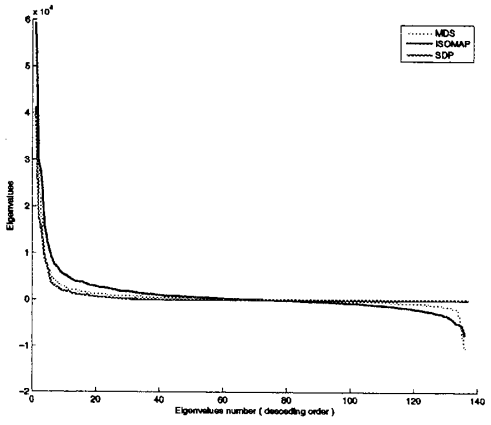

図 1　固有値を降順に並び替えその絶対値の log を出力

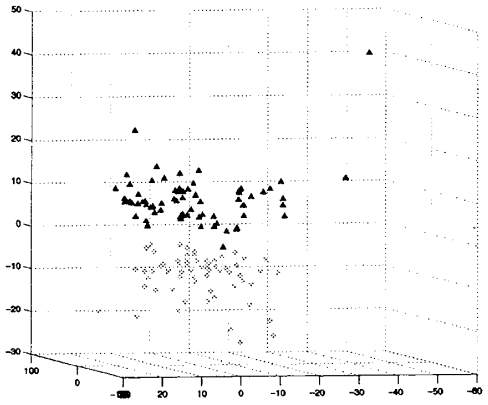

図 2　提案手法による 3 次元埋め込み
青色：クラス' go '，緑色：クラス' please '

$$Error = \frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij}(K))^2}{\sum_i \sum_j d_{ij}^2} \tag{11}$$

$$\hat{d}_{ij}(K) = K(i,i) + K(j,j) - 2K(i,j) \tag{12}$$

手法	提案手法	MDS	ISOMAP
go vs please	0.0016	0.6392	0.6022
sad vs what	0.0014	0.6381	0.6139

表 1　埋め込み前後の距離の誤差

表 1 において，提案手法の誤り率が他の手法より良い結果であると言える．その理由として，提案手法はカーネル行列を固有値解析するときに，負の固有値を出なくするからだと考えられる．さらに，提案手法による 3 D プロットを図 2 に示す．

5.2　実験 2

分類実験では，提案手法とともに，ラプラシアン固有マップ法 (Laplacian Eigenmap，以下 LE 法と略す)，k-近傍法 ($k = 1, 3, 5$) の 3 手法を比較する．LE 法は近傍関係から導かれた類似度行列に基づき，埋め込みを行う手法であり，近傍関係を保存するように埋め込む点では提案手法，および，ISOMAP と同じであるが，カーネル行列を求めない点が提案手法，MDS, ISOMAP と異なっている．埋め込みに際して，距離の非線形変換を行っているので，距離は定量的には保持されないが，分類問題では良い結果を示している [2]．k-近傍法は，同じ距離ベースの手法であり同じ DTW 距離を使うものの，埋め込みを行わない手法である．

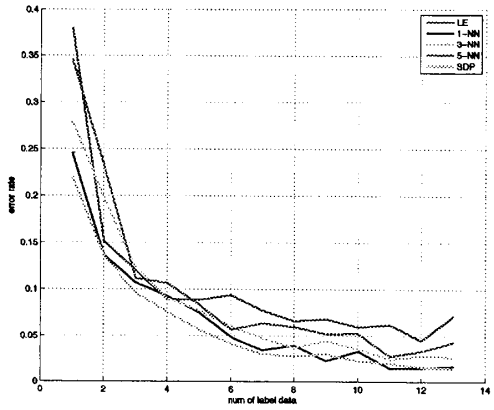

図 3　go vs please 分類結果

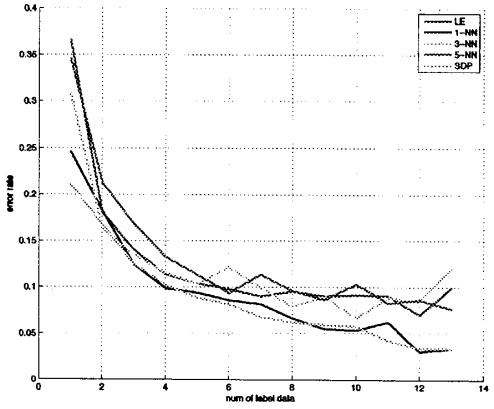

図 4　sad vs what 分類結果

オーストラリア手話言語のラベル付きデータ数を $10, 20, \ldots, 130$ 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．埋め込む次元数は $p = 10$ とし，他のパラメータは実験 1 と同じとした．

図 3，図 4 に，「go」と「please」，「sad」と「what」の分類結果を示す．同じ埋め込み手法の LE 法と比べどちらの分類問題においても提案手法が上回っている事がわかる．さらに，k− 近傍法と比較しても，良好な結果が得られている事がわかる．

5.3　実験 3

実験 3 では，実験 2 で k-近傍法のなかで最も精度が良かった 1-近傍法と提案手法により，新規データの分類実験を行った．テストデータは実験 2 と同じとし，オーストラリア手話言語を用いた．提案手法のパラメータは実験 2 と同じものとする．それぞれ 30 回分類実験を行い平均分類誤り率を見た．

図 5 に，分類結果を示す．新規データの分類実験では，テスト

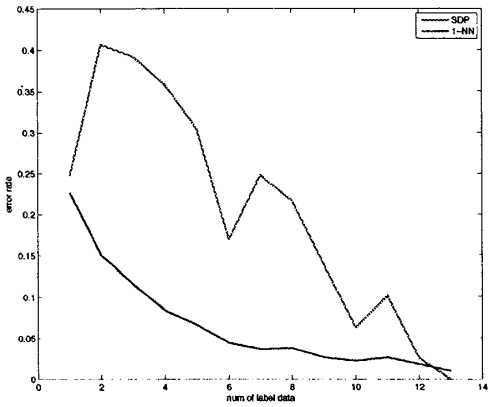

図 5　新規データ分類実験

データが増加するにつれて分類精度が良くなるものの，他の両手法を上回る結果を得る事ができなかった．しかし，テストデータが 130 の時に，両手法よりも良好な結果を得ているので，今後データの規模を増やした場合に両手法を上回る結果が期待できると考えられる．

6　まとめ

本研究では，時間軸のズレに頑強な DTW 距離を用い，半正定値性を満たすカーネル行列を最適化で求め埋め込み，可視化する方法．そして，埋め込み空間において分類する手法を提案した．そして，実データを用いて，埋め込み実験，埋め込み空間における分類実験を行い，その有効性を示した．
今後の研究課題として，新規時系列データの分類実験における精度の向上，サポートベクターマシーン [9] などの，より複雑な分類機を用いる分類実験などが挙げられる．

参考文献

[1] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman & Hall, 2001.
[2] 水原悠子．林朗．末松伸朗．“DTW 距離を用いた時系列データのベクトル空間埋込”，電子情報通信学会論文誌，Vol.J88-D-*II*，No.2，pp.241-249，2005.
[3] L. Rabiner and B. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.
[4] 田村明久．村松正和．最適化法．共立出版株式会社，2002
[5] F.Lu,*et al.*, “Framework for kernel regularization with application to protein clustering”, PANS, Vol.102, No35, 2005
[6] J B. tenenbaum, *et al.*, A global geometric framework for nonlinear dimensionality refuction. *SCIENCE, 290:2319-2323, 2000*
[7] M. Belkin, P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation", Neural Computation, 15 (6),pp.1373-1396, 2003.
[8] S. Hettich et al., "UCI Repository of KDD Databases," http://kdd.ics.uci.edu/, 1999.
[9] C.Corres and V.Vapnik, "Support vector networks", Mach. Learn., vol.20, pp.273-297,1995