

---

DTW 擬距離に基づく  
時系列データのユークリッド空間への埋め込み

---

(18500116)

平成18年度～平成19年度科学研究費補助金（基盤研究C）  
研究成果報告書

平成20年3月

研究代表者 林 朗  
(広島市立大学大学院情報科学研究科教授)

## はしがき

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。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。

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)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。

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検討した。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。サンプル時系列データ集合のカーネル行列を求める問題を、近傍保存埋め込み(NPE)とよぶ半正定値計画問題として定式化した。さらに、新たな時系列データが与えられたとき、NPEで求めたカーネル行列を拡張する問題をサンプル外拡張(OSE)とよぶ半正定値計画問題として定式化した。NPE、OSEにより求めたDTWカーネルの有効性をSVMを用いた分類問題、類似検索に関する予備実験にて検証した。その結果、分類問題においては、k近傍法や、半正定値性を満たさないカーネルを用いた手法より、高い精度が得られ

## 研究組織

- ・研究代表者：林 朗 (広島市立大学大学院情報科学研究科教授)
- ・研究分担者：末松 伸朗 (広島市立大学大学院情報科学研究科准教授)

## 交付決定額 (配分額)

(金額単位：円)

	直接経費	間接経費	合計
平成18年度	1,300千円	0	1,300千円
平成19年度	1,600千円	480千円	2,080千円
総計	2,900千円	480千円	3,380千円

## 研究業績

澤村康匡・林朗・末松伸朗, ナイストロム法を用いた時系列データの高速類似検索法の検討, 第5回情報科学技術フォーラム予稿集, pp. 51-54, 2006

成田浩之 林朗 末松伸朗 岩田一貴, 半正定値計画法による時系列データの埋め込み, 第6回情報科学技術フォーラム予稿集, pp.5-8, 2007

杉浦玄朗, 林 朗, 後藤直人, 末松伸朗, 岩田一貴, 階層隠れ CRF の提案, 第6回情報科学技術フォーラム予稿集, pp. 3-4, 2007

Hiroyuki Narita, Yasumasa Sawamura, Akira Hayashi, Learning a kernel matrix for time series data, Proc. ICONIP 2007 14th International Conference on Neural Information Processing, 2007

Takaaki Sugiura, Naoto Gotou, Akira Hayashi, Hierarchical Hidden Conditional Random Fields, 第21回 信号処理シンポジウム 予稿集, pp. 628-633, 2007

Takaaki Sugiura, Naoto Gotou, Akira Hayashi, A Discriminative Model Corresponding to Hierarchical HMMs, Proc. 8th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'07), pp.375-384, 2007

Takaaki Sugiura, Hierarchical Hidden Conditional Random Fields, Master's Thesis, Graduate School of Information Sciences, Hiroshima City University, 2008

澤村康匡, 動的時間伸縮距離を用いた時系列データの低次元空間への埋め込み, 修士論文, 広島市立大学大学院情報科学研究科. 2008

Hiroyuki Narita, Yasumasa Sawamura, Akira Hayashi, DTW-Distance Based Kernel for Time Series Data, submitted for publication, 2008