

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25870811

研究課題名(和文)再生核ヒルベルト空間の構造最適化による非線形識別に関する研究

研究課題名(英文)A Study on nonlinear classification methods based on the optimization of the structure of the reproducing kernel Hilbert space

研究代表者

日野 英逸 (HINO, Hideitsu)

筑波大学・システム情報系・助教

研究者番号：10580079

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：機械学習の主要な手法の一つとして、カーネル法と呼ばれる手法がある。その効果はカーネル関数の種類とそのパラメタに大きく依存する。変換される確率変数の分布に着目してカーネル関数を最適化する手法の開発を目的とし、以下の3つの成果を得た。特定の判別器を用いた時に最適な判別性能が得られるように、特徴空間のデータ分布最適化の観点からカーネル関数を設計する方法を提案した。次に、データの有する情報量の観点から既存データと新規観測データの類似性の評価尺度を提案し、それに基づくデータの分類手法や外れ値検出手法を提案した。応用として、話者認識やジェスチャ認識、タンパク質の構造判別、変化点検知等の問題に取り組んだ。

研究成果の概要(英文)：Kernel method is one of the most important methods in machine learning. Its effectiveness depends on the kernel function and its parameter. We developed methods for optimizing the kernel function based on the data distribution of the intrinsic high dimensional space associated with the kernel function. We also proposed a similarity measure for the given data and the newly observed data based on the notion of information content in the given data, and applied the proposed measure for classifying a set of data and finding outliers from the observed dataset. We applied the developed methods to speaker recognition, hand gesture recognition, protein structure classification, and change point detection from time series data.

研究分野：機械学習と応用統計

キーワード：機械学習 応用統計 パターン認識

1. 研究開始当初の背景

統計的データ解析において、カーネル法と呼ばれる多変量解析の手法群が盛んに研究されている。判別問題において、非線形関数を用いてデータを高次元特徴空間へ写像することで、既存の線形判別器を用いて高精度な非線形判別システムの実現が期待できる。データの特徴量そのものではなく、データが分布する特徴空間における内積が計算できれば多くの解析アルゴリズムを実行することが可能である。この性質に基づき、カーネル法ではカーネル関数と呼ばれる関数を用いて間接的に特徴空間における内積を計算することで、効率的に非線形データ解析を実行する。カーネル法の効果は、利用するカーネル関数の種類とカーネル関数を特徴付けるパラメタの選択に大きく依存するため、その最適選択がカーネル法の適用における最重要問題である。特に、複数のカーネル関数の凸結合が新たに特徴空間を定義することから、複数の候補カーネル関数を用意し、その凸結合の係数を何らかの基準で最適化する多重カーネル学習と呼ばれる手法が盛んに研究されていた。

2. 研究の目的

本研究では、カーネル関数によってデータが写像される特徴空間が規定されることと、判別器によってはデータが特定の分布をしている場合に最良の性能を発揮することが理論的に明らかになっていることに着目し、データの特徴空間における分布を調整する方法論として多重カーネル学習を捉え、従来の経験リスク最小化とは異なるカーネル関数最適化の枠組みを構築することに取り組む。本研究で目指すカーネル最適化の枠組みの概念図を図1に示す。

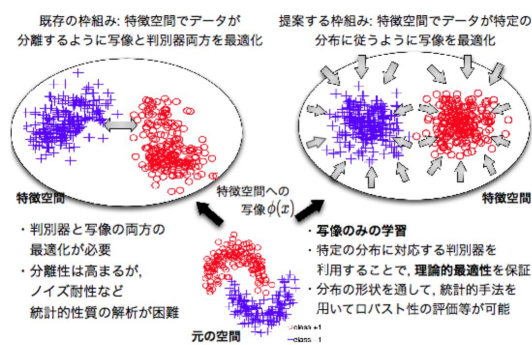


図1: 本研究で目指す枠組みの概念図

特徴空間におけるデータ分布最適化という新しい観点に基づくカーネル関数最適化の方法論を構築し、その理論的性質を解明する。また、音声認識を始めとする諸問題へ応用し、その実用性を検証する。

3. 研究の方法

特徴空間におけるデータ分布を議論する

ためには、特徴空間におけるデータの経験分布と理論的分布との距離尺度として、理論的に妥当かつ、実用上扱いやすいものを選定することが重要な課題となる。分布の形状の検定のための統計量は既に多数提案されており、原理的には特徴空間においてこうした検定統計量を計算し、パラメタ最適化によりその統計量を最適化することで、特徴空間におけるデータ分布を操作することが可能であると考えられる。

しかし、本研究で対象とする特徴空間は非常に高次元になりうるのに対し、従来の検定統計量の多くは1次元データの分布に対してしか適用できない。さらに、従来の検定統計量の多くは計算コストが大きく、また、パラメタが複雑に入り組んだ形で含まれてしまい、カーネル関数の最適化が困難になるという問題がある。そこで、評価及び最適化が容易な分布間の距離尺度として、まずは一般の分布に適用できる経験特性関数を利用する。特性関数は、数学的には確率密度関数のフーリエ変換に相当し、2つの特性関数が一致すればそのフーリエ逆変換である確率密度関数も一致するため、分布が同一であることがいえる。特性関数は、観測したデータを用いて経験特性関数として評価することが可能である。さらに、経験特性関数はデータの内積のみを用いて計算出来るため、データの特徴空間における内積を表すカーネル関数のみを用いて計算できる。つまり、特徴空間における分布の評価を、元の空間のデータをカーネル関数に代入したものを介して自然に行うことが出来るという特長がある。また、経験特性関数はカーネル関数の結合係数に関して可微分かつ有界という、最適化対象として扱いやすい性質を有する。特にデータ分布が正規分布であるかどうかのみを判定する場合には、経験特性関数は非常に簡単な式で書き表すことが出来る。こうした性質に着目し、本研究では分布間の距離尺度として経験特性関数の利用を検討する。

4. 研究成果

(1) 特徴空間におけるデータ分布の観点からのカーネル関数最適化に先立ち、特徴空間においてデータのばらつきを特徴づける量であるエントロピー及びデータの重要性の尺度である情報量の推定手法を複数提案した。観測したデータそれぞれに異なる重みがついている場合の情報量の推定手法や、その応用として非線形情報論的クラスタリング手法を提案し、特にクラスタリング手法においては国際会議にて best paper award を受賞した。

(2) カーネル関数の組合せ最適化方法としては、当初の予定通り、特に判別分析への応用を念頭に置き、判別分析が最も効果的に働くようにデータ分布を修正するようにカー

ネル関数を設計する方法を開発した。これまで研究代表者が提案していた、経験特性関数に基づく手法とは異なり、情報理論における基礎的な不等式である指数エントロピー不等式をカーネル関数最適化の文脈に初めて導入し、効率的なカーネル関数設計の方法論を提案した。国際会議にて本研究成果を発表し、その内容は雑誌論文としても取り上げられた。

(3) 音声による話者識別、及びタンパク質の構造分類のためのタンパク質間類似度尺度の最適化に、提案する方法論を適用し、それぞれの分野で既存手法と同等以上の性能を発揮することを確認した。

5. 主な発表論文等

[雑誌論文](計9件)

Protein Fold Classification using Large Margin Combination of Distance Metrics; Chendra Hadi Suryanto, Kazuhiro Fukui, Hideitsu Hino IEICE Transactions on Information and Systems, Vol.E99-D, No.3, pp.-, March 2016(査読あり)

mmpp: a Package for Calculating Similarity and Distance Metrics for Simple and Marked Temporal Point Process; Hideitsu Hino, Ken Takano, Noboru Murata The R Journal, Vol. 7, Issue 2, pp. 237--248, December 2015(査読あり)

Change-Point Detection in a Sequence of Bags-of-Data; Kensuke Koshijima, Hideitsu Hino, Noboru Murata IEEE Transactions on Knowledge and Data Engineering, vol. 27, No. 10, pp.2632--2644, October 2015, (doi:[10.1109/TKDE.2015.2426693](https://doi.org/10.1109/TKDE.2015.2426693)) (査読あり)

Non-Parametric Entropy Estimators Based on Simple Linear Regression; Hideitsu Hino, Kensuke Koshijima, Noboru Murata Computational Statistics and Data Analysis, Volume 89, September 2015, Pages 72-84, (doi:[10.1016/j.csda.2015.03.011](https://doi.org/10.1016/j.csda.2015.03.011)) (査読あり)

Multi-Frame Image Super Resolution Based on Sparse Coding; Toshiyuki Kato, Hideitsu Hino, Noboru Murata Neural Networks, Vol. 66, pp. 64--78, 2015, (doi:[10.1007/s11265-014-0899-7](https://doi.org/10.1007/s11265-014-0899-7)) (査読あり)

Entropy Power Inequality for Learning Optimal Combination of Kernel Functions; Hideitsu Hino Journal of Signal Processing Systems, Volume 79, Issue 2, pp 201-210, May 2015, (doi:[10.1007/s11265-014-0899-7](https://doi.org/10.1007/s11265-014-0899-7)) (査読あり)

あり)

Intrinsic Graph Structure Estimation Using Graph Laplacian; Atsushi Noda, Hideitsu Hino, Masami Tatsuno, Shotaro Akaho, Noboru Murata Neural Computation, vol. 26, issue 7, Jul, pp. 1455-1483, 2014, (doi:[10.1162/NECO_a_00603](https://doi.org/10.1162/NECO_a_00603)) (査読あり)

A Non-parametric Clustering Algorithm With A Quantile-based Likelihood Estimator; Hideitsu Hino, Noboru Murata Neural Computation, vol. 26, issue 9, September, pp. 2074-2101, 2014, (doi:[10.1162/NECO_a_00628](https://doi.org/10.1162/NECO_a_00628)) (査読あり)

Information Estimators for Weighted Observations; Hideitsu Hino, Noboru Murata Neural Networks, Volume 46, October 2013, Pages 260-275, (doi:[j.neunet.2013.06.005](https://doi.org/j.neunet.2013.06.005)) (査読あり)

[学会発表](計10件)

Spotting Finger Spelled Words from Sign Language Video by Temporally Regularized Canonical Component Analysis; Shohei Tanaka, Hideitsu Hino, Kazuhiro Fukui The IEEE International Conference on Identity, Security and Behavior Analysis 2016, Tohoku University (宮城県仙台市), February 29-- March 2, 2016(査読あり)

Personal Authentication based on 3D Configuration of Micro-Feature Points on Facial Surface; Takao Yoshinuma, Hideitsu Hino and Kazuhiro Fukui The 7th Pacific-Rim Symposium on Image and Video Technology, Auckland, New Zealand, November 23-27 2015(査読あり)

A kernel method to extract common features based on mutual information; Takamitsu Araki, Hideitsu Hino, Shotaro Akaho The 21st International Conference on Neural Information Processing (ICONIP2014), Kuching, Malaysia, November 3-6, 2014(査読あり)

A Non-Parametric Maximum Entropy Clustering; Hideitsu Hino, Noboru Murata International Conference on Artificial Neural Networks 2014(ICANN2014), Hamburg, Germany, September 15-19, 2014(査読あり)

An Algorithm for Directed Graph Estimation; Hideitsu Hino, Atsushi Noda, Masami Tatsuno, Shotaro Akaho, Noboru Murata International Conference on Artificial Neural Networks 2014(ICANN2014), Hamburg, Germany, September 15-19, 2014(査読あり)

Pairwise Similarity for Line Extraction From Distorted Images; Hideitsu Hino, Jun Fujiki, Shotaro Akaho, Yoshihiko Mochizuki, Noboru Murata The 15th International Conference on Computer Analysis of Images and Patterns (CAIP2013), York, UK, August 27-29 2013(査読あり)

Gaussian Multiple Kernel Learning with Entropy Power Inequality; Hideitsu Hino The 2013 IEEE Machine Learning for Signal Processing Workshop (MLSP2013), Southampton, UK, September 22-25 2013(査読あり)

Supervised Covariance Selection For Linear Discriminant Analysis; Hideitsu Hino, Nima Reyhani The 2013 IEEE Machine Learning for Signal Processing Workshop (MLSP2013), Southampton, UK, September 22-25 2013(査読あり)

Integration of MKL-based and i-vector-based speaker verification by short utterances; Hideitsu Hino, Tetsuji Ogawa The 2nd IAPR Asian Conference on Pattern Recognition (ACPR2013), Loisir Hotel & SPA Tower Naha (沖縄県那覇市), November 5-8 2013(査読あり)

Combination of Multiple Distance Measures for Protein Fold Classification; Chendra Hadi Suryanto, Hideitsu Hino, Kazuhiro Fukui The 2nd IAPR Asian Conference on Pattern Recognition (ACPR2013), Loisir Hotel & SPA Tower Naha (沖縄県那覇市), November 5-8 2013(査読あり)

6 . 研究組織

(1)研究代表者

日野 英逸 (HINO, Hideitsu)

筑波大学・システム情報系・助教

研究者番号 : 10580079