統計数理研究所 75 周年記念事業（2019）

正定値カーネルによる統計的機械学習

Statistical Machine Learning by Positive Definite Kernels

数理・推論研究系 福水 健次 （Kenji Fukumizu）

要　　旨

正定値カーネルとそれが定める再生核ヒルベルト空間を用いて，確率分布を表現する方法論を確立し，さまざまな統計的問題に適用してきた．その一連の研究に関して概説する．

キーワード：正定値カーネル，再生核ヒルベルト空間，統計的推論，検定，ベイズ推定

1. はじめに

カーネル法は，データを（非線形）写像することによってデータの高次モーメントを扱う方法論であり，サポートベクターマシンの提案以来，機械学習の主要技術の一つとして発展してきた（福水, 2010）．データに変換を施してから解析する手法は古くから存在するが，カーネル法の特徴は，特殊な内積を持つ関数空間への写像を用いることにより，写像後のデータに対する線形の処理が効率的に行える点にある．

正定値カーネルとは，集合 Ω（データが存在する空間） 上に定義された対称な 2 変数関数 $k:\Omega\times\Omega\to\mathbb{R}$ で，任意の点 $x_1,\dots,x_n\in\Omega$ に対しグラム行列 $(k(x_i,x_j))$ が半正定値性を満たすものである．Ω 上の正定値カーネル k に対し，Ω 上の関数からなるヒルベルト空間 H が定まり，カーネル法ではこれを「特徴空間」と呼ぶ．このヒルベルト空間は特別な内積を有しており，第 2 変数を $x\in\Omega$ に固定して第 1 変数に関する関数とみなした $k(\cdot,x)$ と任意の関数 $f\in H$ の内積が，$\langle f,k(\cdot,x)\rangle_H=f(x)$ と関数値に一致する．この性質を再生性といい，ヒルベルト空間 H を再生核ヒルベルト空間と呼ぶ．

データ解析に正定値カーネルを用いる際には，データの存在する空間 Ω に正定値カーネル k を定め，次の「特徴写像」によって特徴ベクトル $\phi(x)\in H$ を仮想的に作成する．

$$\phi:\Omega\to H,\qquad x\mapsto k(\cdot,x).$$

再生性を用いると，

$$\langle\phi(x),\phi(y)\rangle_H=k(x,y)$$

が得られるが，これは 2 つの特徴ベクトル $\phi(x),\phi(y)$ のヒルベルト空間 H における内積が，正定値カーネルの値の評価によって容易に計算されることを意味しており，カーネル法の鍵となる．データ $x_1\dots,x_n$ に対して特徴ベクトル $\phi(x_1),\dots,\phi(x_n)$ を想定し，これらに線形回帰，主成分分析など様々な既存のデータ解析手法を適用したものがカーネル法として総称されている．

2. 分布の表現とその応用

Ω 上の確率分布 P に対し $\mu_P:=\int k(\cdot,x)dP(x)\in H$ という関数を，分布 P の表現として

用いることが可能である．これを**カーネル平均**と呼ぶ．k が非線形カーネルの場合，カーネル平均は分布 P の高次モーメントの情報を有している．特にガウスカーネルなどは，分布 P から再生核ヒルベルト空間 H への１対１写像 $P \mapsto \mu_P \in H$ を定めることが知られており，確率分布の特性関数と同様の役割を持つ．このようなカーネルを特性的カーネルと呼ぶ（Fukumizu et al., 2004）．確率 P に従う有限個の i.i.d. データ $X_1, \ldots, X_n$ が与えられた場合，μ_P を $(1/n)\sum_{i=1}^{n} k(\cdot, X_i)$ により推定することが可能である．

同時分布 P を持つ確率変数 $(X.Y)$ に対して（非心）**共分散作用素**を $C_{YX} = E_P[k_X(\cdot, X)k_Y(\cdot, Y)]$ により定める．ここで，k_X, k_Y はそれぞれ X, Y が値をとる空間に定義された正定値カーネルである．C_{YX} は有限次元確率ベクトルの共分散行列の一般化であり，X と Y の統計的関係を表現している．これも，グラム行列を用いて有限サンプルから容易に推定可能である．

近年，筆者を含むグループらにより，カーネル平均と共分散作用素を様々な統計的問題へ適用する研究がなされてきた．例えば，Fukumizu et al.（2004)では，特性的なカーネルを導入し，確率変数の条件付き独立性を共分散作用素によって特徴づけ，それを回帰問題における次元削減に応用している．また，２つの分布の距離をカーネル平均の距離により定義し，２標本問題に応用する研究や，独立性，条件付き独立性の検定にカーネル法を応用する研究などが発展している．これらカーネル平均を用いた方法に関しては Song et al.（2013)や Muandet et al.（2017)などを見ていただくとよい．

また，Fukumizu et al.（2013)は，カーネル平均と共分散作用素を用いてベイズ事後確率を推定する方法を提案した．この方法は状態空間モデルにおけるフィルタリングの問題などに応用されている．

カーネル平均や共分散作用素を用いたデータ解析は一般的なノンパラメトリック推論の方法論であり，最近では，近似ベイズ計算への応用，因果推論への応用，さらに深層学習と組み合わせた生成モデルへの応用など，大きな広がりを見せている．本稿によって興味を持たれた方がさらに研究を発展させていくことを期待している

参　考　文　献

福水健次 (2010) 『カーネル法入門 – 正定値カーネルによるデータ解析』, 朝倉書店, 東京.

Fukumizu, K., Bach, F.R. and Jordan, M. I.（2004）. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces, *Journal of Machine Learning Research*, 5, 73–99.

Song, L., Gretton, A., and Fukumizu, K. (2013) Kernel Embeddings of Conditional Distributions. *IEEE Signal Processing Magazine* 30(4), pp. 98–111.

Muandet, K., Fukumizu, K., Sriperumbudur, B. and Sch́olkopf, B. (2017), Kernel Mean Embedding of Distributions: A Review and Beyond, *Foundations and Trends in Machine Learning*, Vol. 10: No. 1–2, pp 1–141.

Fukumizu, K., Song, L., and Gretton, A. (2013) Kernel Bayes' Rule: Bayesian Inference with Positive Definite Kernels. *Journal of Machine Learning Research*, 14, 3753–4783.