

平成21年6月19日現在

研究種目： 基盤研究 (C)
研究期間： 2006~2008
課題番号： 18500184
研究課題名 (和文) 機械学習の手法による非線形複雑系の動的解析とシミュレーション科学への応用
研究課題名 (英文) Dynamical Analysis of Nonlinear Complex Systems by Machine Learning and Applications in Simulation Science
研究代表者
伊庭 幸人 (IBA YUKITO)
統計数理研究所 モデリング系 准教授
研究者番号： 30213200

研究成果の概要：

多変量の時系列データを機械学習の手法である「再生カーネル法」を用いて解析する手法を研究した。系列の背後にマルコフ連鎖に従う確率的なダイナミクスを仮定したとき、カーネル関数を含む一般固有値問題を解くことで、特別な仮定によらずに、背後のマルコフ連鎖の重要な固有値と固有関数を抽出することが可能なことを示した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	900,000	0	900,000
2007年度	800,000	240,000	1,040,000
2008年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,500,000	480,000	2,980,000

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング(B)

キーワード：カーネル法, 機械学習, 非線形, ダイナミクス, シミュレーション

1. 研究開始当初の背景

生体高分子や複雑系物理の諸分野においては、大規模な計算機シミュレーションによって、大量のデータが日々生産されている。たんぱく質・ペプチドやランダム磁性体、非線形媒質等の数値実験による研究はその典型である。

そこでしばしば起こるのは、シミュレーションはできたものの、その結果の分析が困難であるという事態である。これまでのように、各分野の専門家が個別に考えた量や指標を数値実験で計算するだけでなく、有効な変数への縮約や射影を自動的にを行い、本質的な情報を抽出する方法が待望されている。

既存の手法としては、主成分分析 (PCA) があるが、線形部分空間への射影に留まっている点、および、時系列としての特性を十分に生かしていない点が不十分と考えられた。

2. 研究の目的

本研究の目的は、機械学習やデータマイニングの分野で生まれた考え方を非線形系のデータ解析に応用することで、動的かつ非線形の解析を実現し、シミュレーション科学への応用をはかることである。

機械学習の諸方法のなかでも、再生カーネル (reproducing kernel) に基づく手法 (カーネル法) は非線形で雑音を含む関係を容易に

扱うことができ、今世紀に入ってから大きな流行となっている。カーネル法は、サポートベクターマシン(SVM)に代表されるような判別学習のパラダイムから生まれたが、今日では判別だけでなく、情報の圧縮や非線形の関係の抽出に有用なことが示されている。この理念をさらに発展させると、既存の線形多変量解析のあらゆる手法が「カーネル化」によって非線形に拡張できることになる。判別学習がパターン認識やゲノム科学で有効なものに対し、これらの情報圧縮・関係抽出系の手法は、物理学やシミュレーション科学においても力を発揮することが期待される。

ここで提案する研究では、上述のニーズと方法論的背景の自然な展開として、情報圧縮・関係抽出系のカーネル法を非線形複雑系の領域に適用し、時間領域のデータ、高次元・大規模データへ応用可能な手法を開発するとともに、シミュレーションの結果の解析への応用を行なう。

3. 研究の方法

(1) 全般に、主として数値実験によって新しい応用および新しい手法の開発と検証を行う。

(2) 最初のステップとして、既存のカーネル法を非線形複雑系に適用する研究を進展させる。

(3) 次に、本研究の目的のために新しいカーネル法を開発し、その妥当性と有効性を数値実験で確かめる。

(4) その他、カーネル法と非線形・時系列データ解析の接点について、研究過程で得られたアイデアを探求する。

4. 研究成果

(1) カーネル法による一般化同期の解析

① 従来から行ってきた「カーネル正準相関分析 (KCCA) による多変量カオス時系列データの非線形相関の検出」の研究を進展させ、結果を論文としてまとめた。

② 2つの非線形振動子が結合しつつカオス状態にあるとき、両者の間の関係として、(a)完全に同期する (b)互いに相関なく運動する、のほかに、(c)お互いの関係そのものがカオス的・非線形的な特性を示す、という可能性があり、一般化同期と呼ばれている。

こうした現象を時系列データから検出するには線形手法では限界があるが、KCCAの考え方を応用することで、鋭敏で信頼性のある検出が可能になることを数値実験で示すことができた。

とくに、データに対してサロゲーション法を適用することで、相関の非線形部分が検出されていることを確認することができた。

(2) 新しいカーネル手法の開発

① 非線形複雑系の動的解析のために、新しいカーネル法を提案し、数値実験によりその妥当性を検証した。提案手法は、できるだけ一般的な仮定のもとで、データから、背後に想定されるマルコフ連鎖の主要な固有値・固有ベクトルを求めることを目的とする。いくつかの固有ベクトルへの時系列の射影を考えることで、データ縮約・低次元表現が可能になる。

② われわれの提案する手法は、高野と宮下によって1995年に統計物理学の分野で提案された「緩和モード解析」をカーネル化することで導出される。いずれの手法でも、多変量時系列の定常性を仮定した上で、(ある時刻の各変量の値、時刻 τ での各変量の値)というサンプルの組が十分沢山得られたとして、これを解析する。

高野・宮下の方法では、この条件のもとで、固有ベクトルが簡単な形でパラメトリックに表現できると仮定し、時間遅れのある相関と同時刻相関を含む一般固有値問題を導出する。主要な固有値・固有ベクトル(混合の遅いモードに対応する固有値・固有ベクトル)はその解として求められる。

これに対し、提案手法では、固有ベクトルに関する仮定をゆるめて、これをカーネル関数によってデータ依存的に表現することで、グラム行列と「時間遅れ」を含んだグラム行列を含む一般固有値問題を導出する。主要な固有値はその一般固有値として、固有ベクトル(混合の遅いモードに対応する固有値・固有ベクトル)は一般固有ベクトルからカーネル表現を通じて、それぞれ求めることができる。この手法はまた、高野・宮下の方法の双対表現を与えているとみなすこともできる。

③ 時間反転に対して統計的性質を保つ時系列(いまの場合、詳細釣り合いを満たすようなマルコフ連鎖によって生成された時系列といってもよい)だけを考えた場合、高野・宮下の方法は、マルコフ連鎖の遷移行列に対応するレイリー商を最大化するという変分原理から、テスト関数をパラメトリックに選ぶことによって導出できることが知られている。

ここでカーネル関数をテスト関数に選ぶことで、別の経路から高野・宮下の方法の「カーネル化」を行うことが可能である。やや意外なことに、この方法で得られる結果は直接にカーネル化を行ったものと厳密には一致しないが、適当な条件のもとではほぼ同等となることが期待される。

④ 提案法から導かれる一般固有値問題は、オリジナルの高野・宮下法のそれとは異なり、

サンプル数×サンプル数のサイズの行列を含む数値的にきわめて悪条件のものとなる。

本研究では、この一般固有値問題を扱う手法を2通り与え、それぞれを数値実験で検証した。はじめに試みた手法は、一般固有値問題に含まれる(時間遅れの無い)グラム行列を固有値分解し、有意にゼロでない固有値に対応する固有ベクトルを抜き出し、両辺をそれらが張る部分空間に射影する、というものであるが、固有値問題を解くための計算量が多くなるという欠点がある。

次に試みた改良手法は、カーネル行列にピボット選択付きのコレスキー分解を適用して、有意にゼロでない固有値に対応する空間を抜き出すもので、一般固有値問題の方程式の両辺を同時に処理することで、最初の方法と等価な結果を高速に得るものである。この方法は、カーネルCCAなどで用いられている手法の応用であるが、これの導入によって、提案手法は現実的なサイズのデータセットに適用可能となった。

⑤ 提案手法は、(1)状態数の少ないマルコフ連鎖のシミュレーションデータ (2)小さなイジングスピン系のシミュレーションデータ (3)エネルギーに極小値が複数あるような連続変数系のメトロポリス法によるシミュレーションデータ、などについてテストされ、妥当な結果を与えることが示された。

(3) ガウス過程を事前分布としたベイズモデルに対するマルコフ連鎖モンテカルロ法の開発

① シミュレーションとカーネル法の接点、というテーマを考えていく途上で「連続時間のガウス過程を事前分布としたベイズモデルの扱い」という問題に逢着し、その研究を行った。

② ガウス過程を事前分布としたベイズモデルの場合、事後分布を最大にする解(MAP解)を求めるためには、サンプル点の上で定義されたカーネルの線形結合の範囲で考えれば良いということが知られている(リプリゼンター定理)。しかしながら、事後分布からのサンプル生成が必要な場合、リプリゼンター定理は使えない。

この問題は一般的には解決不能と思われるが、問題を限定して、事後分布からのサンプル生成が特定の目的のためであるとすれば、解決することが可能な場合があることがわかり、研究中である。開発中の方法では、カーネル行列にピボット選択付きのコレスキー分解を適用する、という別の項目で述べた手法が重要な役割を果たす。

③ 提案手法は非線形関係や非線形時系列の

解析に有用であると考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2件)

①S. Akaho and K. Fukumizu
Active Learning for Network Estimation
Proc. of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Mar. 31- Apr. 5, Hawaii, 2007, 402-409

②H. Suetani, Y. Iba, K. Aihara
Detecting generalized synchronization between chaotic signals:
a kernel-based approach
Journal of Physics A: Mathematical and General, 2006, 10723-10742

[学会発表] (計 4件)

①伊庭幸人, Kernel Eigenmode Analysis of time series, Advanced Monte Carlo Research group 1st Workshop, 2009年6月9日, 統計数理研究所(東京)

②伊庭幸人, 緩和モード解析の改良
日本物理学会大会 64 回年次大会, 2009年3月30日, 立教大学(東京)

③伊庭幸人, カーネル法による緩和モード解析,
日本物理学会第 63 回年次大会, 2008年3月26日, 近畿大学(東大阪)

④伊庭幸人, カーネル法によるマルコフ連鎖の遷移行列の固有値解析, 統計数理研究所研究会 非線形科学と統計科学の対話, 2007年11月28日, 統計数理研究所(東京)

[図書] (計 1件)

赤穂昭太郎, カーネル多変量解析 (「確率と情報の科学」シリーズ) 岩波書店, 2008年11月。

6. 研究組織

(1) 研究代表者

伊庭 幸人 (IBA YUKITO)
統計数理研究所・モデリング研究系・准教授
研究者番号: 30213200

(2) 研究分担者

赤穂 昭太郎 (SHOTARO AKAHO)
産業技術総合研究所・脳神経情報研究部門・

情報数理研究グループ長
研究者番号：40356340
(2006年度)

(3)連携研究者
赤穂 昭太郎 (SHOTARO AKAHO)
産業技術総合研究所・脳神経情報研究部門・
情報数理研究グループ長
研究者番号：40356340
(2007～2008年度)