

研究種目： 基盤研究 (C)
 研究期間： 2007～2009
 課題番号： 19500249
 研究課題名 (和文)
 グラフを用いた統計的学習と計算推論に関する幾何学的研究とその応用
 研究課題名 (英文)
 Geometric Study on Statistical Learning and Computation based on Graphs
 研究代表者
 福水 健次 (FUKUMIZU KENJI)
 統計数理研究所・モデリング研究系・教授
 研究者番号： 60311362

研究成果の概要 (和文)： 本研究では、複雑な構造を持ったデータを扱う際に重要となる、グラフ構造によって記述されるデータや確率分布に対して、その推論や計算に関する理論的研究を行なった。その結果、(1) ネットワーク構造推定の新しい能動的学習法、(2) 無限次元指数分布族の導入による柔軟な推論の枠組み、(3) グラフ構造を用いた確率推論アルゴリズムである確率伝搬法の性質の解明に向けた新しい数理的方法、の提案をそれぞれ行い、従来では困難であった推論課題の解決や理論的性質の解明を導いた。

研究成果の概要 (英文)： Theoretical study was carried out on the inference and computation of data and probabilities described by graph structures, which is of importance in analyzing data of complex structure. This study proposed (1) a method of active learning of network structure, (2) a framework of flexible inference by introducing infinite dimensional exponential families, and (3) a mathematical method for analyzing theoretical properties of the belief propagation, which is an inference algorithm with graph structure. Each of these methods provided solutions to inference problems and elucidation of theoretical properties, which had been difficult by conventional methods.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	270,000	1,170,000
2008年度	1,600,000	480,000	2,080,000
2009年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野： 統計科学

科研費の分科・細目： 情報学・統計科学

キーワード： グラフ, 統計的推論, 確率伝搬法, 情報幾何学, 位相幾何学, カーネル法, 再生核ヒルベルト空間

1. 研究開始当初の背景

近年の計算機科学や生命科学の発達は、相互に参照しあうネットワーク上の文書や、遺

伝情報に関する大規模データといった、複雑な構造を持った高次元・大容量データの出現を促し、データを扱う統計科学や情報処理の

分野では、単なるベクトルではない、グラフやネットワークなどの複雑な構造を持つデータを処理する方法の開発がきわめて重要となっている。グラフ構造を扱うデータ解析に関する重要な研究課題には、(1) データからのグラフ構造の学習、(2) 学習によって獲得されたグラフを用いた統計的推論、(3) それらの処理に必要な確率計算の効率的方法、などが含まれる。

研究開始当初の背景を述べる。(1) のグラフ構造の学習については、グラフィカルモデルの構造学習など、古くから研究が行われているが、近年では、バイオインフォマティクスにおける蛋白質ネットワークの推定のように、非常に多くの変数間の関係を抽出するニーズが高まっており、確率分布に基づく厳密な方法に限らず、さまざまな方法を試みる必要がある。従って、ネットワーク構造推定に対して高精度で簡便なアルゴリズムを提供することは、重要な課題のひとつとなっていた。

(2) で述べたグラフを用いた統計的推論に関わる課題のひとつに、変数間の相関関係がグラフやツリーで与えられた際に、その情報を関連する他のベクトル/非ベクトルデータとどのように統合して用いるかという問題がある。これに対する有効な方法論のひとつに、正定値カーネルによるカーネル法がある。分担者の赤穂と代表者の福水は、データの依存関係をカーネル法によって解析する研究を精力的に行ってきた。さらに、福水はカーネル法と無限次元指数分布族の結びつきを明らかにした。この観点を発展させ、無限次元指数分布族と無限次元情報幾何によってカーネル法の理論的研究を行うことは興味深い課題であった。

(3) に挙げた確率計算の効率的方法としては、確率伝搬法が広く研究されている。このアルゴリズムは、グラフによって与えられた変数間の条件付独立性を利用して、事後確率などの計算を効率的に行うものである。グラフにループがある場合には、近似的に確率計算を行うループ付確率伝搬法がよく用いられるが、収束性や近似可能性といった理論的性質は必ずしも明らかとなっておらず、特別な場合の正当性や緩い十分条件が与えられている程度であり、今後の研究が待たれている状態であった。

2. 研究の目的

1 で述べた背景に基づき、本研究は次の3つの主課題を設定した。

課題 A: ネットワークの能動的学習

課題 B: カーネル法の無限次元情報幾何的研究

課題 C: 確率伝搬法の位相幾何的研究

課題 A は、ネットワーク推定の問題に対する能動的学習ないしは最適実験計画を用いた方法論の確立と、その効率的な計算方法に関する研究である。能動的学習は、より効率的な推定・学習を目的としてデータ採取法を最適化する方法論であり、研究代表者の福水がニューラルネットに適用した研究などがある。福水と赤穂らは、ネットワーク推定に対する能動的学習法の基本的提案を行いその効果を確認したが、そのアルゴリズムはさまざまな近似を必要としていた。本研究では、ネットワーク構造に関連する情報幾何的な考察を行うことによって、これら従来の近似手法を理論的に理解し改良していくことを目的とする。

課題 B では、有限次元の指数分布族に対する様々な統計手法を無限次元の関数空間に拡張する研究を行い、その手法の性質を無限次元情報幾何を用いて明らかにすることを目的とする。

課題 C は、グラフ上の確率伝搬法のアルゴリズムを、グラフの位相的性質を用いて解明しようとする。確率伝搬法のアルゴリズムの性質に、そのグラフの位相的性質が強く反映することは容易に想像されるが、その関係を陽に記述する研究は今まで見られなかった。本研究では、代数的位相幾何の定式化を用いることにより、グラフの持つ位相的普遍量と確率伝搬法のアルゴリズムの性質との理論的関係を明らかにすることを目的とする。

3. 研究の方法

研究開始である 19 年度は、まず理論的な考察を中心に研究を進める。並行して 19 年度後半において、アルゴリズム検証のための計算機実験を行うプラットフォームを PC ワークステーション上に構築する。比較的小規模なデータによる実験を中心に実験を行い、理論の検証と問題点を考察し、以降の研究の深化を目指した。

各課題において研究する具体的問題は以下を予定した。

課題 A

- ・ 定常状態データを用いたネットワークの能動学習を研究する。
- ・ ネットワークの能動学習における事後分布の近似手法の開発
- ・ 事後分布の近似手法の情報幾何的な解析
- ・ 非系列データへの拡張
- ・ 遺伝子ネットワークシミュレータを用いた、アルゴリズムの有効性検証

課題 B

- ・ 正定値カーネルによる無限次元指数分布族(カーネル指数分布族)の理論的基盤の整備
- ・ カーネル法を用いた、ノンパラメトリックなグラフィカルモデル構造学習の研究。

- ・ カーネル指数分布族を用いた、その他のさまざまな統計的手法の拡張

課題 C

- ・ 代数的位相幾何の観点からの解析
- ・ 複数のループを持つグラフにおけるループ付確率伝搬法の位相幾何的解析
- ・ ループ付確率伝搬法の収束性とグラフの位相的普遍量との関係の研究
- ・ 期待値伝搬法の情報幾何的・位相幾何的研究

研究の海外協力者として、津田宏治、Arthur Gretton, Bernhard Schoelkopf などと共同で行うとともに、世界の最新の情報を入手して研究の方向性にフィードバックした。

4. 研究成果

まず、課題 A については、平衡状態の値からネットワークを推定する方法において、情報量にもとづく能動学習法の基本的なアルゴリズムを提案し、遺伝子ネットワークのシミュレーションデータによる基本的な性能確認を行った。提案手法はネットワーク学習に対して、計算の簡便な最適設計基準を与えている。課題 A に関しては、本研究期間内では、実データへの適用などの展開を図ることはできなかったが、基本的なアルゴリズムの有効性を確認しており、今後の発展の基礎を与えた。

次に課題 B に関しては、正定値カーネルの定める無限次元指数分布族の情報幾何的研究を行い、一般には理論的困難が伴うことが知られている無限次元指数分布族の接ベクトル空間上への双対接続の導入に関して、導入が可能であるための十分条件を導き、その条件を満たすモデルの例を示した。

また、有限次元の指数分布族には埋め込めないが、無限次元指数分布族の部分モデルとして実現可能な有限混合分布族を考え、また、特異点の近傍での最尤推定量の漸近的性質に関する新しいアプローチを提案した。この方法は特異点に関する従来のアプローチよりもはるかに単純な幾何学的方法を提供している。

本研究で得られたこれらの結果は、無限次元情報幾何と正定値カーネル法の研究を結びつける新しい方向性を与えている。結果の一部は書籍 Algebraic and Geometric methods in statistics の一編として出版されている。

課題 C に関しては、大きく分けて3つの成果が得られた。第1は、統計物理学や確率伝搬法で重要な分配関数の計算手法に関して考察を行い、確率伝搬アルゴリズムが与える周辺分布の近似値と真の周辺分布との差を、

グラフ上の一般化ループによって展開する新しい方法を提案し、それに基づいて確率伝搬法の近似精度に関する理論的考察を行った。これは、ループがある場合の確率伝搬アルゴリズムの理論解析に新しい方法論を与えるものである。この結果をまとめた論文は Journal of Physics A に掲載されている。

第2の成果は、グラフ多項式に係る理論的結果である。第1の成果で得られたループ展開式がグラフ多項式と密接に関係していることを発見し、新たなグラフ多項式の提案とその理論的性質の研究を行った。現状では、確率伝搬法や統計物理的な研究との関連は明確とは言えないが、このグラフ多項式の性質の解明が進めば、確率伝搬アルゴリズムの理論的性質の解明が進展すると期待される。

第3の成果は、確率伝搬法と密接に関連することが知られているベータ自由エネルギー関数の幾何学的研究である。今まで、確率伝搬法の不動点とベータ自由エネルギーの停留点一致することが知られていた。本研究では、ベータ自由エネルギーのヘッセ行列式と、グラフ理論や整数論で知られるグラフゼータ関数の多変量拡張を結びつける関係式を証明し、その関係式を用いて確率伝搬法の理論的性質を解明した。グラフゼータ関数はグラフの様々な位相的情報を有するため、この関係式は確率伝搬法の数学的解析に対して新しい方法論を与えるという意義を持つ。確率伝搬法に関して新しく得られた理論的結果として、第1に、ベータ自由エネルギーのヘッセ行列の正定値性条件を導き、ベータ自由エネルギーが凸関数であることとグラフが2個以上の独立なサイクルを持たないことが同値であることを示した。第2に、確率伝搬法の固定点の局所安定性の条件を明らかにした。第3に、確率伝搬法の固定点が一意になるための新しい十分条件を導き、サイクルを2個持つある種のグラフィカルモデルに対する確率伝搬法の解が一意であることを示した。これらは、すべて既存の解析手法では容易に導くことのできなかった性質であり、グラフゼータ関数の方法が有力な道具となることを示している。また、グラフゼータ関数の保有する情報をさらに詳しく調べることによって、確率伝搬法の幾何学的な研究の将来の発展に本質的な役割を担うことが期待される。

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 10 件)

- ① Sriperumbudur, B., K. Fukumizu, A. Gretton, G. Lanckriet, B. Scholkopf.

- Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions. *Advances in Neural Information Processing Systems*. 査読有, 22, 2010, to appear.
- ② Gretton, A., Z. Harchaoui, K. Fukumizu, B. Sriperumbudur. A Fast, Consistent Kernel Two-Sample Test. *Advances in Neural Information Processing Systems* 査読有, 22, to appear (2010)
- ③ Watanabe, Y. and K. Fukumizu. Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation. *Advances in Neural Information Processing Systems*. 査読有 22, to appear (2010)
- ④ Watanabe, Y. and K. Fukumizu. New graph polynomials from the Bethe approximation of the Ising partition function. *Combinatorics, Probability and Computing*. 査読有, 2010, to appear.
- ⑤ B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Scholkopf, G.R.G. Lanckriet. Hilbert Space Embeddings and Metrics on Probability Measures. *Journal of Machine Learning Research*. 査読有, 2010, 11, 1517-1561.
- ⑥ Y. Watanabe, K. Fukumizu. Loop series expansion with propagation diagrams. *Journal of Physics A: Mathematical and Theoretical*. 査読有. 42(4), 2009. オンライン 045001.
- ⑦ K. Fukumizu. Exponential manifold by reproducing kernel Hilbert spaces. *Algebraic and Geometric methods in statistics*. 査読有. 2009. 291-306.
- ⑧ K. Fukumizu, B. Sriperumbudur, A. Gretton, B. Schoelkopf. Characteristic Kernels on Groups and Semigroups. *Advances in Neural Information Processing Systems*. 査読有, 22, 2009, 473-480
- ⑨ K. Fukumizu, F.R. Bach and M. Jordan. Kernel dimension reduction in regression. *The Annals of Statistics*. 査読有, 37(4), 2009, pp. 1871-1905.
- ⑩ M. Cuturi, K. Fukumizu. Kernels on structured objects through nested histograms. *Advances in Neural Information Processing Systems*, 19. 査読有. 2007, 329-336

[学会発表] (計 21 件)

- ① 渡辺有祐, 福水健次. グラフ上のメッセージ伝搬アルゴリズムに現れるグラフゼータ関数について. 2010年度日本数学会年会. 2010年3月25日 慶応義塾大学.
- ② 福水健次. グラフ上の効率的計算アルゴリズムと幾何学. 科研費研究会: 新たな幾何学探訪勉強会. 2010年3月16日. 松島
- ③ 渡辺有祐, 福水健次. ベーテ自由エネルギーと Loopy belief Propagation に現れるグラフのゼータ関数について. 第12回情報論的学習理論ワークショップ (IBIS 2009) 2009年10月19日. 九州大学
- ④ 福水健次. Nonparametric Tests of Conditional Independence with Reproducing Kernels. Dagstuhl Seminar 09401: Machine learning approaches to statistical dependences and causality. 2009年9月28日. ドイツ (Dagstuhl)
- ⑤ 福水健次. Conditional independence and dimension reduction with positive definite kernels. The 1st Institute of Mathematical Statistics AsiaPacific Rim Meeting. 2009年7月1日. 韓国(ソウル)
- ⑥ 福水健次. Dependence Analysis with Positive Definite Kernels. 理研滞在型研究会: 階層性を持つ大自由度力学系の時系列解析 2009年5月24日 理化学研究所 (和光)
- ⑦ 福水健次. Independence and Conditional Independence with Reproducing Kernels. BIRS Workshop on Understanding the New Statistics. 2008. 9. 16. Banff, Canada
- ⑧ 福水健次. 再生核ヒルベルト空間を用いた統計的推論. 京都大学数理解析研究所共同利用研究集会. 2008. 9. 10. 京都大学数理解析研究所.
- ⑨ 福水健次. Dependence Analysis with Reproducing Kernel Hilbert Spaces. 7th World Congress on Statistics and Probability. 2008. 7. 16. シンガポール.
- ⑩ 福水健次. Painless embeddings of distributions: the function space view. International Conference on Machine Learning 2008 (チュートリアル) 2008. 7.5. Helsinki, Finland
- 福水健次. Reproducing Kernel Exponential Manifold: Estimation and Geometry. *Mathematical Explorations in Contemporary*

- Statistics. 2008. 5. 20. Sestri Levante, Italy.
- ⑪ 福水健次. Independence, conditional independence, and characteristic kernels. German-Israel Foundation Workshop. 2008.5.16. Tuebingen, Germany.
- ⑫ 福水健次. 再生核ヒルベルト空間による指数分布族と統計的推定. 大阪市立大学数学研究所ミニスクール「情報幾何への入門と応用 II」2007年12月22日. 大阪
- ⑬ 福水健次. カーネル法による条件付独立性尺度とその因果推論への応用. 計算世界観ワークショップ. 2007年12月13日. 東工大
- ⑭ 福水健次. Kernel Measures of Conditional Dependence. 21st Annual Conference on Neural Information Processing Systems. 2007年12月5日 バンクーバー(カナダ)
- ⑮ A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Scholkopf, A. Smola. A Kernel Statistical Test of Independence. 21st Annual Conf. on Neural Information Processing Systems. 2007年12月4日. バンクーバー(カナダ)
- ⑯ 白石友一, 福水健次. 多値判別における二値判別機の組み合わせ法の学習について. 第10回情報論的学習理論ワークショップ (IBIS 2007) 2007年11月5日. 横浜
- ⑰ 渡辺有祐, 福水健次. ループ展開による分配関数の解析と近似. 第10回情報論的学習理論ワークショップ (IBIS 2007) 2007年11月5日. 横浜
- ⑱ 福水健次. Learning Causal Structure with Kernel-based Dependence Measures. International Workshop on the Interface between Statistical Causal Inference and Bayesian Networks. 2007年11月4日横浜.
- ⑲ 福水健次. 再生核ヒルベルト空間を用いた統計的推論. 実解析学シンポジウム 2007 (科研費研究集会) 2007年10月19日. 大阪
- ⑳ X. Sun, D. Janzing, B. Scholkopf and K. Fukumizu. A kernel-based causal learning algorithm. The 24th Annual International Conference on Machine Learning. 2007年6月23日 オレゴン (アメリカ合衆国)

[図書] (計0件)

[産業財産権]
○出願状況 (計0件)

○取得状況 (計0件)

[その他]
ホームページ
<http://www.ism.ac.jp/~fukumizu/>

6. 研究組織

(1) 研究代表者

福水 健次 (FUKUMIZU KENJI)
統計数理研究所・モデリング研究系・教授
研究者番号: 60311362

(2) 研究分担者

池田 思朗 (IKEDA SHIRO)
統計数理研究所・数理・推論研究系・
准教授
研究者番号: 30336101

赤穂 昭太郎 (AKAHO SHOTARO)
産業技術総合研究所・脳神経情報研究
部門・グループ長
研究者番号: 40356340

(3) 連携研究者

()

研究者番号: