

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 20 日現在

機関番号：11301

研究種目：基盤研究（B）

研究期間：2010～2012

課題番号：22300078

研究課題名（和文） 量子力学的ラベリングにより拡張されたデータマイニング理論の創出

研究課題名（英文） Creative extensions of data mining theory by means of quantum-mechanical labeling

研究代表者

田中 和之（TANAKA KAZUYUKI）

東北大学・大学院情報科学研究科・教授

研究者番号：80217017

研究成果の概要（和文）：

量子力学的に拡張されたマルコフ確率場によるベイジアンネットワークシステムのデータマイニングへの応用のための理論体系の基盤の構築を行った。量子確率伝搬法の定式化を確立し、量子統計的学習理論に道を開くことに成功したことは特出すべき成果である。また、一次相転移を伴うシステムに対する統計的学習方式を拘束条件付きエントロピー最大化の視点から構築することに成功したことは想定されなかった成果である。

研究成果の概要（英文）：

General frameworks of statistical machine learning systems based on quantum-mechanical extended Markov random fields has been provided by using belief propagation methods. Some novel statistical model selection schemes by using quantum belief propagation methods have been proposed. As one of unexpected research results in our present project proposal, we have established new statistical learning frameworks for some statistical models with the first-order phase transitions in the stand point of constrained maximization of entropies instead of conventional maximum likelihood estimations.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	4,900,000	1,470,000	6,370,000
2011 年度	4,600,000	1,380,000	5,980,000
2012 年度	4,300,000	1,290,000	5,590,000
年度			
年度			
総計	13,800,000	4,140,000	17,940,000

研究分野：情報学

科研費の分科・細目：人間情報学・ソフトコンピューティング

キーワード：数理工学、機械学習、統計数学、情報統計力学、確率的情報処理

1. 研究開始当初の背景

量子力学的に拡張された状態を情報処理に応用する研究は、量子計算、量子誤り訂正符号、量子暗号など、次の時代の情報通信技術としてこれまで多くの学術的研究が行われている。しかしながら実際に情報通信技術に

応用する段階となると、はたして量子コンピュータは実現可能なのか？量子通信はどこまで現実のものとなるのか？など未知数の部分が多いともいうことができる。そのような中で、現在のコンピュータのなかで、量子情報の理論的な枠組みを活用しようという

試みが、近年、徐々に始まりつつある。そのひとつに量子アニーリングという方法がある。この手法は最適化問題におけるシミュレートッドアニーリングの熱ゆらぎを量子ゆらぎに置き換えようというものであり、田中和之(本課題研究代表者)の「画像処理への応用を目的とする量子統計力学的反復計算法の提案」(信学論 (A), Vol. J80-A, 2117, 1997)という形で世界に先駆けて発展したものである。最近では Santoro 氏等の研究グループ(Trieste, Italy)が量子アニーリングの最適化問題に対する有効性を主張する論文(Science, Vol. 295, 2427, 2002)を発表するなど世界的に学術的研究の流れが形成され、代表者の研究成果が世界的に先駆的研究として評価されている。しかし、この量子アニーリングは従来の情報処理の問題を解くアルゴリズムのひとつとして提案されたものであり、問題設定の段階から本質の意味において量子力学的状態を情報処理の状態として用いたものであるわけではない。そこで本課題研究代表者は、これまでの量子アニーリングなどにおいて得られた研究成果をふまえながら、この量子力学的重ね合わせにより形成された状態を情報処理における問題設定の段階において導入し得る研究課題としてデータマイニングに着目し、データのラベル付けにおけるラベルの状態に対して量子力学的重ね合わせによる状態を導入するという着想に至った。

2. 研究の目的

本研究は量子力学的状態を考慮した量子ベイジアンネットワークシステムのデータマイニングへの応用のための理論体系の基盤の構築を行うことを目的としている。例えば、新聞の記事を分類するという問題設定を考えた場合、その記事が「政治」、「経済」、「社会」、「スポーツ」などのいずれかに完全に分類できるかという点必ずしも完全な分類を仕切れない場合の方が多い。むしろ、通常は、そのいくつかのラベル付けが同時にされる方が自然であることがよく起こりうる。そのような状況を、系統的に解決する理論を量子力学という物理学の伝統的概念と、統計的学習理論という近年注目されつつある計算技法を融合することで構築する。例えば YouTube, Google 等に代表されるような Web 上のデータマイニング汎用性のある新しいキーアプリケーションを創出することが期待される。

3. 研究の方法

本研究では隠れマルコフモデルによる統計的学習アルゴリズムの量子力学的拡張をまず行う。その上でより複雑なグラフ構造をもつ量子力学的確率モデルから生成された

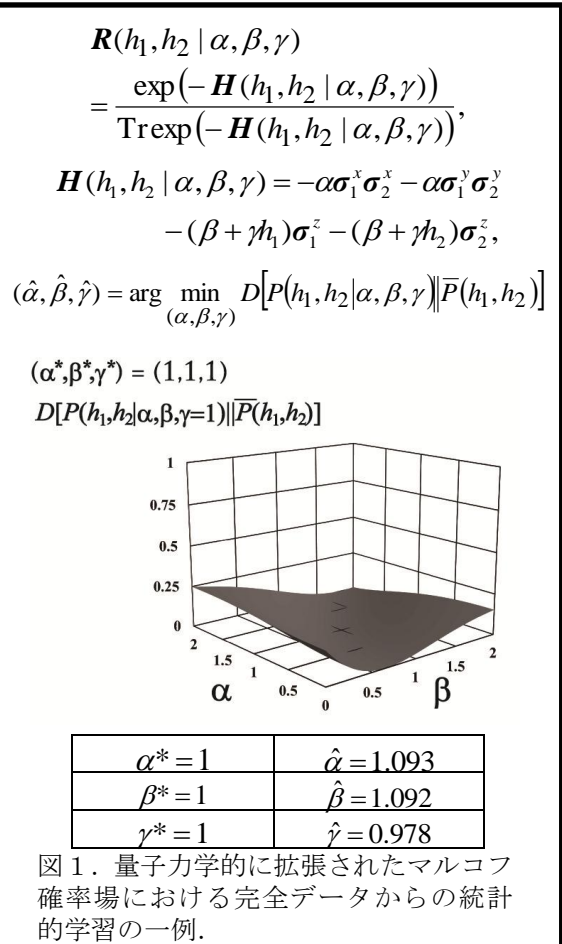
データのラベル付けという視点において進めてゆく。具体的なアルゴリズムは代表者がすでに学術論文 K. Tanaka: Journal of Physics: Conference Series, Vol. 143, Article No. 012023, 2009 のなかで提案している一般化された量子確率伝搬法を理論的基盤として適用することで実現してゆく。数値実験による性能評価はサンプリングを行うために量子力学的に拡張されたマルコフ連鎖モンテカルロ法を再構築した上で進める。その再構築は量子モンテカルロ法とよばれる統計力学の手法を統計的学習理論に組み合わせることで実現する。同時に情報統計力学的計算手法を用いた典型性能評価法の理論的基盤を整備してゆく。

4. 研究成果

(1) 2010 年度は統計的性能評価手法をより一般化されたマルコフ確率場に拡張するアルゴリズムの開発実装を行った。期待値最大化(Expectation Maximization: EM)アルゴリズムによる統計的機械学習アルゴリズムの実装と試験的数値計算を海外研究協力者である D. M. Titterton 教授(英国グラスゴー大学)との研究打合せを通しての助言を受けながら進めた。その成果の一部は平成 22 年 12 月にアドコムメディア社から出版された「——CVIM チュートリアルシリーズ—— コンピュータビジョン最先端ガイド 3」のなかで「第 6 章. 大規模確率場と確率的画像処理の深化と展開」を分担執筆し、その中で公開している。更に拡張された量子マルコフ確率場からデータを生成する量子モンテカルロ法の定式化を行い、具体的にデータを生成するアルゴリズムの開発・実装を連携研究者津田宏治主任研究員(産業技術総合研究所)からの助言を受けながらすすめて、生成されたデータのヒストグラムからモデルの妥当性を確認するための数値実験の準備をすすめた。その一方で、研究分担者の井上純一准教授(北海道大学)は量子確率場モデルにおけるデータ生成過程の統計的解析手法の定式化に成功した。

(2) 2011 年度はマルコフ確率場を密度行列の表現による量子マルコフ確率場への拡張の定式化と拡張された量子マルコフ確率場からデータを生成する量子モンテカルロ法の定式化を具現化するプログラムを実装し、具体的なデータの生成に関する数値実験をすすめて、生成されたデータのヒストグラムからモデルの妥当性の確認を行った。また、これと平行して、研究分担者の井上純一准教授(北海道大学)からの助言をうけながら、スピングラス理論における平均場モデルに対する統計的解析手法の量子力学的に拡張されたマルコフ確率場による統計的機械学習理論への転用を部分的に行った。図 1 はその成

果の一部である。量子力学的自由度をもつデータから、生成モデルのハイパラメータの良好な推定を行うことができることを示している。更に、安田宗樹助教（東北大学大学院情報科学研究科）との協力により、量子確率伝搬法による統計的学習理論の構築に向けての準備の一つとして、スピングラス理論の 1RSB と TAP 方程式の立場からの確率伝搬法による学習アルゴリズムの構築と離散型のチェビシェフ多項式を用いた一般の離散状態をもつマルコフ確率場モデルにおける確率伝搬法による統計的機械学習理論の構築を行った。これらの定式化の一部は確率的画像処理計算モデル、都市交通の予測モデルへと応用し、その有効性の検証を行った。これらの成果の一部は「Philosophical Magazine: The Study of Condensed Matter」, 「Journal of the Physical Society Japan」などを中心とする世界的にインパクトファクターの高い学術雑誌の査読付き原著論文として公開している。



(3) 2012 年度は密度行列の表現による量子力学的に拡張されたマルコフ確率場に対する確率伝搬法のアルゴリズム設計とプログラム実装を行い、モデルの妥当性の確認を行っ

た。また、生成されたデータからの統計的学習理論に従うモデル選択および予測・推論を行う学習アルゴリズムに対する確率伝搬法を用いた定式化を進め、具体的なデータマイニングの問題に適用可能であることを確認した。また、マルコフ確率場を用いたラベリングに共通して内在する問題として一次相転移を伴う確率モデルにおける確率伝搬法と EM アルゴリズムによる統計的学習アルゴリズムを設計する新しい概念を発見するに至った。そのアプローチは拘束条件付きエントロピー最大化を出発点とするものであり、これにより、従来は尤度のモデルパラメータに対する微分化の可能性を仮定し、期待値最大化アルゴリズムを用いていたところが、1次相転移が生じる場合には尤度が相転移点で微分不可能となるため、期待値最大化アルゴリズムを直接適用することが難しいという問題があった。拘束条件付きエントロピー最大化という視点を出発点とすることにより、この問題を解決し、従来法を含む改良された期待値最大化アルゴリズムという形で統計的学習システム設計理論を構築することが可能となった。この着想は海外研究協力者である Lenka Zdeborova 研究員 (CEA Saclay, フランス) との研究打ち合わせを通して得られたものであり、当該研究計画で当初想定されなかった成果のひとつである。

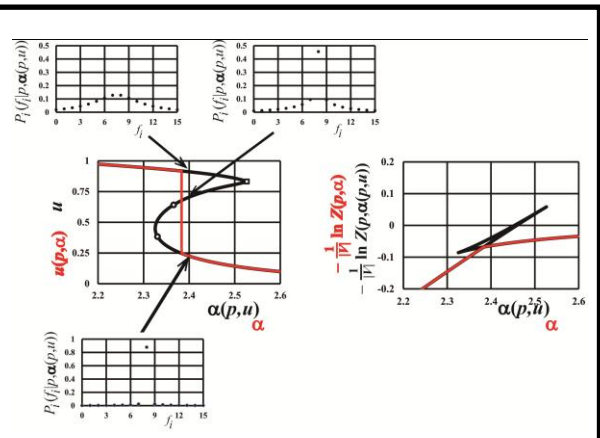


図 2. 拘束条件付きエントロピー最大化の視点により確率伝搬法による統計的学習システムを再構築した数値実験例。黒の実線が提案手法。赤の実線が従来の最尤法による手法。提案手法ではモデルパラメータの決定に重要な役割を果たす (α, u) 曲線が連続曲線となっていることがわかる。

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 14 件)

- ① Muneki Yasuda and Kazuyuki Tanaka: Susceptibility Propagation by Using Diagonal Consistency, Physical Review E, 査読有り, Vol.87, No.1, Article No.012134, pp.1-6, January 2013, DOI:10.1103/PhysRevE.87.012134.
- ② Muneki Yasuda, Junya Tannai and Kazuyuki Tanaka: Learning Algorithm for Boltzmann Machines Using Max-Product Algorithm and Pseudo-Likelihood, Interdisciplinary Information Sciences, 査読有り, Vol.18, No.1, pp.55-63, December 2012, DOI:10.4036/iis.2012.55.
- ③ Kazuyuki Tanaka, Muneki Yasuda and D. Michael Titterton: Bayesian Image Modeling by Means of Generalized Sparse Prior and Loopy Belief Propagation, Journal of the Physical Society of Japan, 査読有り, Vol.81, No.11, Article No.114802, pp.1-9, November 2012, DOI: 10.1143/JPSJ.81.114802.
- ④ Yosuke Otsubo, Jun-ichi Inoue, Kenji Nagata and Masato Okada: Effect of Quantum Fluctuation in Error-Correcting Codes, Physical Review E, 査読有り, Vol.86, Article No.051138, pp.1-10, November 2012, DOI: 10.1103/PhysRevE.86.051138.
- ⑤ Muneki Yasuda, Yoshiyuki Kabashima and Kazuyuki Tanaka: Replica Plefka Expansion of Ising Systems, Journal of Statistical Mechanics: Theory and Experiment, 査読有り, Vol.2012, No.4, P04002, pp.1-16, April 2012, DOI: 10.1088/1742-5468/2012/04/P04002.
- ⑥ Muneki Yasuda, Shun Kataoka and Kazuyuki Tanaka: Inverse Problem in Pairwise Markov Random Fields using Loopy Belief Propagation, Journal of the Physical Society of Japan, 査読有り, Vol.81, No.4, Article No.044801, pp.1-8, April 2012, DOI: 10.1143/JPSJ.81.044801.
- ⑦ Shun Kataoka, Muneki Yasuda and Kazuyuki Tanaka: Statistical Analysis of Gaussian Image Inpainting Problems, Journal of the Physical Society of Japan, 査読有り, Vol.81, No.2, Article No.025001, pp.1-2, February 2012, DOI: 10.1143/JPSJ.81.025001.
- ⑧ Muneki Yasuda and Kazuyuki Tanaka: TAP Equation for Nonnegative Boltzmann Machine: Philosophical Magazine: The Study of Condensed Matter, 査読有り, Vol.92, Nos.1-3, pp.192-209, January 2012, DOI:10.1080/14786435.2011.634856.
- ⑨ Shun Kataoka, Muneki Yasuda, Kazuyuki Tanaka and D. Michael Titterton: Statistical Analysis of the Expectation-Maximization Algorithm with Loopy Belief Propagation in Bayesian Image Modeling, Philosophical Magazine: The Study of Condensed Matter, 査読有り, Vol.92, Nos.1-3, pp.50-63, January 2012, DOI:10.1080/14786435.2011.624558.
- ⑩ Jun-ichi Inoue: Pattern-recalling processes in quantum Hopfield networks far from saturation, Journal of Physics: Conference Series, 査読有り, Vol.297, Article No.012012, pp.1-10, May 2011, DOI:10.1088/1742-6596/297/1/012012.
- ⑪ Muneki Yasuda, Tetsuharu Sakurai and Kazuyuki Tanaka: Learning Algorithm in Restricted Boltzmann Machines using Kullback-Leibler Importance Estimation Procedure (Invited Paper), Nonlinear Theory and Its Applications, IEICE, 査読有り, Vol.2, No.2, pp.153-164, April 2011, DOI: 10.1587/nolta.2.153.
- ⑫ 安田宗樹, 田中和之: 相関等式を用いたボルツマンマシンの決定論的近似学習アルゴリズム, 電子情報通信学会論文誌(D), 査読有り, Vol.J93-D, No.11, pp.2446-2453, November 2010.
- ⑬ Kazuyuki Tanaka, Shun Kataoka and Muneki Yasuda: Statistical performance analysis by loopy belief propagation in Bayesian image modeling, Journal of Physics: Conference Series, 査読有り, Vol.233, Article No.012013, pp.1-11, July 2010, DOI:10.1088/1742-6596/233/1/012013.
- ⑭ Jun-ichi Inoue: Deterministic flows of Order Parameters in Stochastic Processes of Quantum Monte Carlo Method, Journal of Physics: Conference Series, 査読有り, Vol.233, Article No.012010, pp.1-14, July 2010, DOI: 10.1088/1742-6596/233/1/012010.

[学会発表] (計 13 件)

- ① 北川智隆, 安田宗樹, 田中和之: ガウシアングラフィカルモデルを用いた確率的画像補修フィルタの設計, 電子情報通信学会ニューロコンピューティング研究会, 査読無し, 2012年11月17日, 東北大学.
- ② 浅利岳, 安田宗樹, 和泉勇治, 田中和之: 2部グラフ型ボルツマンマシンに対する複合最尤法, 電子情報通信学会ニューロコンピューティング研究会, 査読無し, 2012年11月17日, 東北大学.
- ③ 田中和之, 安田宗樹: スパース相互作用をもつ事前分布と確率伝搬法によるベイジアンモデリング, 日本物理学会 2012年秋季大会, 査読無し, 2012年9月18日, 横浜国立大学.
- ④ Kazuyuki Tanaka: Bayesian image modeling by generalized sparse Markov random fields and loopy belief propagation, Interdisciplinary Workshop on Inference ---Information processing in complex systems with applications to traffic forecasting---, 査読無し, 12 June, 2012, Inria Paris-Rocquencourt (Paris sub-office), Paris, France.
- ⑤ Muneki Yasuda, Shun Kataoka, Yuji Waizumi and Kazuyuki Tanaka: Composite Likelihood Estimation for Restricted Boltzmann Machines, 21st International Conference on Pattern Recognition, 査読有り, 14 November, 2012, Tsukuba, Japan.
- ⑥ 片岡駿, 安田宗樹, 樺島祥介, 田中和之: 圧縮センシングを用いた画像補修モデリング, 日本物理学会第67回年次大会, 査読無し, 2012年3月27日, 関西学院大学西宮上ヶ原キャンパス.
- ⑦ 村山竜太, 安田宗樹, 和泉勇治, 田中和之: カラーチャンネル間の相関を考慮した Gaussian FoE モデル, 電子情報通信学会情報論的学習理論と機械学習研究会, 査読無し, 2011年11月9日, 奈良女子大学.
- ⑧ Marvin Huang, Chiou-Ting Hsu and Kazuyuki Tanaka: Automatic Ridge Network Detection in Crumpled Paper Based on Graph Density, 2011 IEEE 13th International Workshop on Multimedia Signal Processing, Digital Object Identifier, 査読有り, 19 October, 2011, Hangzhou, China.
- ⑨ 田中和之: スパース正則化項をとものうマルコフ確率場における確率伝搬法, 電子情報通信学会ニューロコンピューティング研究会, 査読無し, 2010年11月18日, 東北大学.
- ⑩ 片岡駿, 安田宗樹, 田中和之: ランダム欠損を利用した画像圧縮法, 日本物理学会 2011年秋季大会, 査読無し, 2011年9月22日, 富山大学.
- ⑪ 安田宗樹, 田中和之: オンサガー項を考慮した感受率伝搬法の改良とその一般化, 日本物理学会 2011年秋季大会, 査読無し, 2011年9月21日, 富山大学).
- ⑫ 安田宗樹, 樺島祥介, 田中和之: 1RSB を考慮した TAP 方程式を用いた感受率伝搬法, 日本物理学会 2010年秋季大会, 査読無し, 2010年9月25日, 大阪府立大学中百舌鳥キャンパス.
- ⑬ Muneki Yasuda, Yoshiyuki Kabashima and Kazuyuki Tanaka: Plefka Expansion of Replica Extended Ising Model and Its Application to Inverse Ising Problem, 24 IUPAP International Conference on Statistical Physics, 査読無し, 22 July, 2010, Cairns, Australia.

[図書] (計 3 件)

- ① 安田宗樹, 片岡駿, 田中和之共著 (分担執筆): ---CVIM チュートリアルシリーズ--- コンピュータビジョン最先端ガイド3(八木康史, 斎藤英雄編), 第6章. 大規模確率場と確率的画像処理の深化と展開, pp.137-179, アドコム・メディア株式会社, December 2010
- ② Jun-ichi Inoue, Yoshiyuki Kabashima, Kazuyuki Tanaka and Toshiyuki Tanaka (eds): International Workshop on Statistical-Mechanical Informatics 2010, Journal of Physics, Conference Series, Vol.233, Article Nos. 011001-011002, Article Nos. 012001-012023, IOP Publishing, July 2010.
- ③ 田中和之, 安田宗樹共著 (分担執筆): 映像情報メディア工学大事典(映像情報メディア学会編)/基礎編, 第4部門. 画像処理, 第12章. 確率的画像処理, pp. 233-235, オーム社, June 2010.

6. 研究組織

(1) 研究代表者

田中 和之 (TANAKA KAZUYUKI)

東北大学・大学院情報科学研究科・教授
研究者番号: 80217017

(2) 研究分担者

井上 純一 (INOUE JUN-ICHI)

北海道大学・大学院情報科学研究科・准教授
研究者番号: 30311658

(3) 連携研究者

津田 宏治 (TSUDA KOJI)
産業技術総合研究所・生命情報工学研究
センター・主任研究員
研究者番号：90357517