

研究種目：若手研究（スタートアップ）

研究期間：2008～2009

課題番号：20800012

研究課題名（和文） 近似ベイズ推論によるデータ解析法の漸近特性

研究課題名（英文） Asymptotic analysis of approximate Bayesian inference methods

研究代表者

渡辺 一帆 (WATANABE KAZUHO)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：10506744

研究成果の概要（和文）：高次元データの次元圧縮とクラスタリングのために、制約つき指数型分布族の混合モデルに対する変分ベイズ法による学習アルゴリズムを導出し、手書き数字認識の実データに応用し手法の有効性を検証した。また、イベントの発生を表す系列から変動する発生率を推定するモデルである変動二項過程に対し、局所変分法による近似推定法を導き、近似、推定精度を実験的に調べた。さらに局所変分法の情報論的な意味づけを行い、一般的な推定法の導出を行った。

研究成果の概要（英文）：We developed an approximate inference method for the constrained exponential-family mixture models that are used for the simultaneous dimensionality reduction and clustering of high-dimensional data. It was applied to the hand-written digit recognition task and its effectiveness was demonstrated. We also derived an approximate inference method for the varying binomial process that efficiently estimates the varying probabilities of some event. Furthermore, we demonstrated the general framework and the information-theoretic view of the local variational approximation.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,220,000	366,000	1,586,000
2009年度	1,120,000	336,000	1,456,000
年度			
年度			
年度			
総計	2,340,000	702,000	3,042,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：ベイズ推定，変分ベイズ法，局所変分近似，混合モデル，変動二項過程，情報幾何，階層ベイズ法

1. 研究開始当初の背景

パターン認識、バイオインフォマティクス、神経科学等の分野において、大量の高次元データから、その背後に潜む真の構造を抽出す

る目的で、様々な学習モデルと共にベイズ学習の近似学習アルゴリズムが用いられている。近似学習アルゴリズムの有効性は多くの実問題で実験的に知られてきているが、近似

精度、予測精度などの性質は解明されてこなかった。このため近似アルゴリズムの設計には問題ごとに工夫がなされるだけで、設計者の試行錯誤が必要とされてきた。このため、近似学習アルゴリズムの近似精度や予測精度を理論的に評価することで、目的に応じて近似法を有効利用するための学習アルゴリズムの最適設計の指針が求められていた。

2. 研究の目的

本研究は、近似学習アルゴリズムの理論的性質を評価し、その上で予測精度と計算量の観点からそれらを有効利用するための設計の指針を与えることを目的とする。近似精度を明らかにすることで、近似学習アルゴリズムにおける損失関数の振る舞いが、学習モデルの大きさ、事前分布の設定にどのように影響されるかを明らかにし、特に学習データ数が大きい状況の解析を行うことで、近似学習アルゴリズムの特徴を理解することを目指す。近似学習アルゴリズムは想定した確率分布の中から真のベイズ事後分布に対し最も近似精度の良い分布を探索するアルゴリズムであるため、近似精度を調べることにより、学習アルゴリズムの挙動を理解することができる。これにより学習の結果構成される近似事後分布を用いた際の予測精度を明らかにし、予測精度に関して最適になるように近似学習アルゴリズムの設計法を与える。

3. 研究の方法

近似学習アルゴリズムの近似精度及び予測精度を、数理統計や情報理論的な方法により解析し、これらの量が近似事後分布のどのような性質によって特徴づけられるかを明らかにする。そして、データ解析法における近似学習アルゴリズムを予測精度に関して最適になるように、また、ある程度の予測精度を保証した上で効率良く計算を行うように改良を行う。次元圧縮やクラスタリングのために用いられる学習モデルについて、変分ベイズ法や期待値伝播法といった既存の学習アルゴリズムを用いた場合の近似精度を解析する。また、近似の良さを測る尺度に関して一般化を行うことにより、既存の学習アルゴリズムを含む一般の近似アルゴリズムについて成り立つ性質を明らかにする。検証のために人工データやパターン認識における実データを用いた計算機シミュレーションを研究の各段階で行う。

4. 研究成果

(1) 高次元データの次元圧縮とクラスタリングの同時最適化のために、指数型分布族の制

約つき混合モデルに対する変分ベイズ法による学習アルゴリズムを導出した。指数型分布を用いることで、データがバイナリ (0 か 1) や回数等の (実数値でない) 特殊な型を持つ場合に、適切な距離尺度を用いたデータ解析が可能である。次元圧縮を組み合わせたクラスタリングのための変分ベイズ法では、通常のクラスタリングのアルゴリズムに加え、さらに計算困難な積分を行う必要がある。本研究ではラプラス近似を用いてこれら近似する手法を構成し、一般の指数型分布に用いることができる学習アルゴリズムの導出に成功している。また、人工データ及び手書き数字認識の実データを用いた数値実験により、手法の有効性を検証し論文誌に発表した。図1は手書き数字認識における例題数 (横軸) に対する本手法の認識率 (縦軸) を示している。実線、破線は、それぞれ次元圧縮を行わない場合及び2次元に圧縮した場合の結果を示しており、○で示した従来法の結果と匹敵する結果が得られることや、例題数が少ないときの次元圧縮の効果が確認できる。

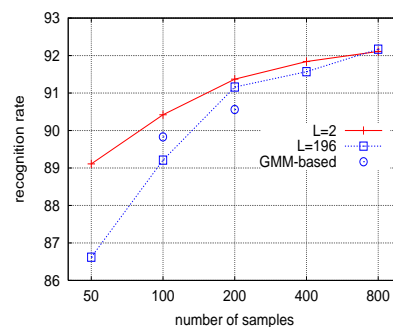


図1 手書き数字認識データの認識率

特殊な型をもつデータの次元圧縮のために主成分分析 (PCA) を拡張した手法として、指数型分布 PCA (e-PCA) が提案されている。本研究で用いた学習モデルは、e-PCA の確率モデルとしての解釈から導出されており、これらのデータ解析手法の統計的推測としての性質を議論することが可能になった。

(2) 平滑化事前分布を用いた変動ポアソン過程及び変動ベルヌーイ過程の推定について、ベイズ推定を厳密に行うための転送行列法、及び事後確率最大化のための逐次最適化法を導出し論文を発表した。これらのモデルは神経科学におけるスパイク系列を含む、あるイベントの発生を表す系列からイベントの発生率を推定するモデルとして用いられている。しかしながら、ベイズ事後分布を解析的に計算することができないため、高速な計算には近似法が有効である。上記の手法は厳

密計算を行うものであり、近似法の評価に用いることができる。実際に、同様のモデルに対する近似推定法を局所変分近似により導出しており、厳密計算との比較による近似精度の評価を行っている。図2に局所変分法による変動生起確率の推定の例を示す。この例では厳密計算を行うのに比べ、数千倍程度の速さで推定を行うことができ、他の近似法に比べて精度の良い近似が実現できることが確認されている。

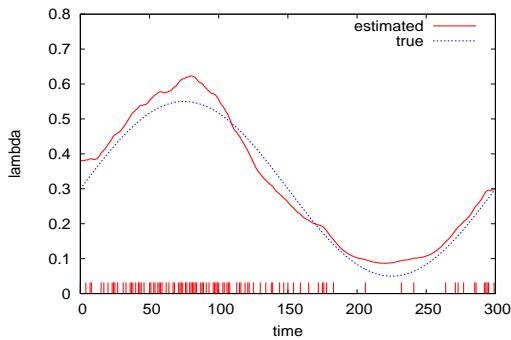


図2 変動生起確率の推定

(3)局所変分近似による事後分布の近似では、周辺尤度の上界及び下界が目的関数となるが、これらを近似事後分布と真の事後分布間の情報量と関係付け、近似事後分布が分布間の情報量を最小化する分布の近似を与えていることを示した。また、上界と下界を結びつける効率的な手法を情報幾何学的考察に基づき構成した。この研究では、局所変分近似による学習アルゴリズムの一般的な枠組みを与えており、これまで各応用においてモデル毎に考えられてきた学習アルゴリズムの導出を、広い視点から記述することに成功している。これにより、新たな応用に対するアルゴリズムの導出が容易になり、局所変分法やその改良について、幾何学的な議論が可能になった。

(4)近赤外光断層撮影法による線形逆問題について、変分ベイズ法を用いた階層ベイズ法の近似解法において事前分布の設定が推定結果に与える影響を調べた。階層ベイズ法では再構成される断層画像において、局所的に変化が見られる領域が存在し、その他大部分では変化が見られないような推定結果を得ることができる（これをスパースな解と呼ぶ）。推定対象のスパース性を変えた人工データを生成し数値実験を行い、事前分布の設定によりスパースな解が得られる場合と、得られない場合があることを確認した。また、その間には推定結果のスパース性に急激な変化が見られることがわかった。これは他の学習モデルについても共通して見られる現

象であり、変分ベイズ法の目的関数である変分自由エネルギーの解析により、理論的な裏付けを行うことは今後の課題である。

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5件)

- ① 赤穂昭太郎, 渡辺一帆, 岡田真人, “指数型分布族の空間におけるデータ解析法について”, 統計数理, 掲載確定, 査読有
- ② 渡辺一帆, 久保智洋, 岡田真人, ”平滑化事前分布を用いた変動2値情報源の事後確率最大化法” 電子情報通信学会論文誌A, J93-A, 326-330, 2010, 査読有
- ③ K.Watanabe, S.Akaho, S.Omachi, M.Okada, ”Variational Bayesian Mixture Model on a Subspace of Exponential Family Distributions”, IEEE Transactions on Neural Networks, vol.20, 1783-1796, 2009, 査読有
- ④ K.Watanabe, H.Tanaka, K.Miura, M.Okada, “Transfer matrix method for instantaneous spike rate estimation”, IEICE Transactions on Information and Systems, vol. E92-D, 1362-1368, 2009, 査読有

〔学会発表〕(計10件)

- ① 渡辺一帆, “局所変分法を用いた近似ベイズ学習における分布間の情報量”, 第54回システム制御情報学会研究発表講演会, 京都市サンプラザ, 平成22年5月19日
- ② 渡辺一帆, “事後分布の局所変分近似における分布間の情報量について”, 電子情報通信学会ニューロコンピューティング研究会, 玉川大学, 平成22年3月10日
- ③ 宮本敦史, “階層変分ベイズ法によるNIRS-DOTの逆問題解法”, 電子情報通信学会ニューロコンピューティング研究会, 玉川大学, 平成22年3月10日
- ④ 渡辺一帆, “近似ベイズ推定とその応用例”, 第24回信号処理シンポジウム, 鹿児島県, 鹿児島市, 平成21年11月26日
- ⑤ K.Watanabe, “Simultaneous Clustering and Dimensionality Reduction Using Variational Bayesian Mixture Model”, 11th IFCS Biennial Conference, Dresden, Germany, 17th, March, 2009.

[その他]

<http://hawaii.naist.jp/~wkazuho/>

6. 研究組織

(1) 研究代表者

渡辺 一帆 (WATANABE KAZUHO)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：10506744

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：