

自己評価報告書

平成 23年 5月 6日現在

機関番号：35302

研究種目：基盤研究(C)

研究期間：2008~2011

課題番号：20500263

研究課題名（和文）グラフィカルモデルにおける統計数値計算法の効率化と加速化に関する研究

研究課題名（英文）Acceleration of statistical iterative algorithms for graphical models

研究代表者

黒田 正博(KURODA MASAHIRO)

岡山理科大学・総合情報学部・准教授

研究者番号：90279042

研究分野：計算統計

科研費の分科・細目：情報学，統計科学

キーワード：EM アルゴリズム，交互最小二乗法，マルコフ連鎖モンテカルロ法，グラフィカルモデル

1. 研究計画の概要

現在進行中の研究課題は以下の3つである。

- (1) ベクターイプシロン法による EM アルゴリズムの加速の理論的研究

EM アルゴリズムおよび拡大 EM アルゴリズムのベクターイプシロン加速化法の収束性の研究

- (2) グラフィカルモデルにおける p -値計算アルゴリズムの研究

マルコフ基底に基づくマルコフ連鎖モンテカルロ法によるグラフィカル対数線形モデルの p -値計算法の研究

- (3) 尺度混在した主成分分析における交互最小二乗法の加速と収束性について

計量データと非計量データが混在した主成分分析で適用される交互最小二乗法の加速化と収束性に関する研究

これらの研究は、EM アルゴリズムのベクターイプシロン加速化法で得た結果を拡張したものである。そして、グラフィカルモデルを統計モデルとして考えた時に、モデル構造を利用して計算効率を図ろうと試みるものである。

2. 研究の進捗状況

上に示した3つの課題それぞれについての進捗状況を記す。

(1)については、グラフィカル対数線形モデル上での EM アルゴリズムおよび拡大 EM アルゴリズム (ECM アルゴリズム) における加速法を提案し、その収束性に関する理論的考察をおこなった。この研究では、ベクターイプシロン法のスカラー版である Aitken デルタ二乗法を用いた。この研究では、パラメータの同時確率分布を条件付き独立分布の積

に変換することで、Aitken デルタ二乗法の適用を可能にした。そして、これまでに知られているこの加速法の収束定理と我々の定式化から、この研究で提案した加速法の収束性を証明した。その成果は、*Statistics and Probability Letters* に掲載された。

現在は、ベクターイプシロン法による ECM および ECME アルゴリズムの加速法を研究している。これの収束定理を与えることができれば、より幅広い統計モデルへの適用が可能になる。

(2)については、4元分割表のグラフィカルモデルにおけるマルコフ連鎖モンテカルロ法による p -値計算法を提案した。この論文の数値実験において、従来の p -値計算で用いられるカイ二乗近似は疎な分割表において当てはまりが非常に悪いという事実を実例で示すことができた。また、4元分割表のグラフィカルモデルにおいて全てのマルコフ基底の move が同じ生成アルゴリズムで記述できることも示した。その成果をまとめたものが、*Computational Statistics* に掲載された。

現在、多次元分割表におけるマルコフ基底の move の生成アルゴリズムへの拡張を考えている。

(3)については、尺度混在した主成分分析における交互二乗法の加速化法を提案し、数値実験において性能評価をおこなった。その成果を、国際会議 (COMPSTAT2010) および *Computational Statistics & Data analysis* に発表した。

3. 現在までの達成度

3つの研究課題については、その成果が専門雑誌に掲載されており、研究目標を概ね順調

に進展していると考えている。

4. 今後の研究の推進方策

「研究の進捗状況」で示した課題の解決を、共同研究者とともに積極的に進めていく。3つの課題については、進行状況はまちまちであり、解決したものから順に成果を発表していく予定である。また、2011年度が本研究の最終年度であるため、研究論文としてまとめる作業もおこなっていくことを考えている。

5. 代表的な研究成果

(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 6 件)

- ① Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M. (2011). Acceleration of the alternating least squares algorithm for principal components analysis. Computational Statistics and Data Analysis, 55, 143-153. [査読有]
- ② Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M. (2010). Improvement of acceleration of the ALS Algorithm using the vector ϵ algorithm. Proceedings of COMPSTAT 2010, e-book, Physica-Verlag HD, 1239-1246. [査読有]
- ③ Kuroda, M., Hashiguchi, H. and Nakagawa, S. (2010). Computing p-values in conditional independence models for a contingency table. Computational Statistics, 25, 57-70. [査読有]
- ④ Kuroda, M., Sakakihara, M. and Geng, Z. (2008). Acceleration of the EM and ECM algorithms using the Aitken δ^2 method for log-linear models with partially classified data. Statistics and Probability Letters, 78, 2332-2338. [査読有]
- ⑤ Wang, M., Kuroda, M., Sakakihara, M. and Geng, Z. (2008). Acceleration of the EM algorithm using the vector epsilon algorithm, Computational Statistics, 23, 469-486. [査読有]

[学会発表] (計 11 件)

- ① Sakakihara, M. and Kuroda, M. Acceleration of convergence for the alternating least squares iteration. 4th CSDA International Conference on Computational and Financial Econometrics. 2011年12月11日. University of London.
- ② 黒田正博, 森裕一, 飯塚誠也, 榊原道夫. 交互最小二乗法の加速化-数値的検討. 統計関連学会連合大会. 2011年9月8日. 早稲田大学.
- ③ 黒田正博, 森裕一, 飯塚誠也, 榊原道夫. 主成分分析における交互最小二乗法の

加速化. 科研費シンポジウム「多変量データ解析法への 計算機統計学・行動計量学的アプローチの新展開」. 2009年12月4日. 大阪大学.

- ④ Kuroda, M., Hashiguchi, H., Nakagawa, S. and Geng, Z. Random table generation for decomposable log-linear models by Markov basis. Workshop on Computational Algebraic Statistics, Theories and Applications. 2008年12月10日. Kyoto University.
- ⑤ Sakakihara, M. and Kuroda, M. Double-secant acceleration for EM and ECM Algorithms for log-linear models with partially classified data. IASC 2008. 2008年12月6日. Pacifico Yokohama.

[その他]

研究成果の HP

<http://www.soci.ous.ac.jp/~kuroda/publication.html>