

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 24 年 6 月 1 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

研究成果の概要（和文）：

vector epsilon 法を組み込んだ EM アルゴリズムの加速法の効率化法を提案し、グラフィカルモデルへの適用を試みた。また、Markov chain Monte Carlo によるグラフィカルモデルの p-値計算のアルゴリズムを開発した。さらに、Alternating Least Squares アルゴリズムに vector epsilon 法を適用した加速法を開発し、数値実験によりその有効性を実証した。

研究成果の概要（英文）：

We developed the acceleration method for the EM algorithm using the vector epsilon algorithm, and then applied it to decomposable log-linear models with missing data. We also proposed MCMC using Markov bases for decomposable log-linear models. Furthermore, we developed a new acceleration method for the Alternating Least Squares (ALS) algorithm using the vector epsilon algorithm, and demonstrated that the acceleration algorithm improves the computational efficiency of the original ALS algorithm.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,200,000	360,000	1,560,000
2009 年度	900,000	270,000	1,170,000
2010 年度	700,000	210,000	910,000
2011 年度	600,000	180,000	780,000
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

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、グラフィカルモデル、交互最小二乗法、マルコフ連鎖モンテカルロ法

1. 研究開始当初の背景

EM アルゴリズムの加速法として vector epsilon 法を accelerator として用いた方法を提案し、収束性と加速性の定理を与えることができた。さらに、提案した EM アルゴリズムの加速法をグラフィカルモデルの構造特性を利用したアルゴリズムに改良し、反復

演算の効率化を図るとともに、収束スピードの加速を目指した。さらに、効率化という観点による“グラフアルゴリズムを用いた局所計算による MCMC の開発”，そして加速という観点から“Alternating least squares (ALS) アルゴリズムの加速化法”の研究に取り組んだ。

2. 研究の目的

研究課題として、統計計算法の効率化を掲げ、次の3つに取り組んだ。

(1) グラフィカルモデルにおける EM アルゴリズムの vector epsilon 加速法の改良

計算の効率化と収束スピードの加速を目指した改良 ϵ -accelerated EM アルゴリズムを開発を目指す。

(2) グラフィカルモデルにおける局所計算を用いた MCMC の改良

Markov chain Monte Carlo (MCMC) によるアプローチを考える。グラフ構造が単純な場合のグラフィカルモデルにおいては、局所計算 (Local computation) を利用した MCMC によるパラメータ推定アルゴリズムをすでに提案した。このアイデアを拡張するとともに、グラフ理論にもとづくグラフアルゴリズムを取り入れた効率化 MCMC を提案し、その妥当性を理論的側面と数値実験により検証する。

(3) Alternating Least Squares (ALS) アルゴリズムの加速化

EM アルゴリズムと同様に ALS アルゴリズムも線形収束するため、収束スピードは遅い。提案した ϵ -accelerated EM アルゴリズムに適用した vector $\cdot \cdot$ アルゴリズムを accelerator とした加速化 ALS アルゴリズムの開発に取り組む。まず、ALS アルゴリズムの収束性を理論的に明らかにし、加速化 ALS アルゴリズムとオリジナルの ALS アルゴリズムとの数値実験による比較をおこなうことで、性能評価をおこなう。また、 ϵ -accelerated EM アルゴリズムと同様に、収束性と加速性の数学的証明を与えることを目指す。

3. 研究の方法

研究代表者の黒田および3名(榊原, 森, 中川)の研究分担者, そして研究目的を確実に達成するため, Geng Zhi 教授 (Peking university) を研究協力者として加え, 本申請研究を遂行した。研究分担者の役割は以下の通りである。

黒田正博

役割: 研究全般

内容: ϵ -accelerated EM アルゴリズムの改良, 局所計算による MCMC および加速化 ALS アルゴリズムの開発

榊原道夫 (岡山理科大学)

役割: グラフィカルモデルでの ϵ -accelerated EM アルゴリズムの改良と加速化 ALS アルゴリズムの開発

内容: 収束スピードの加速についての理論的評価

中川重和 (倉敷芸術科学大学)

役割: 局所計算による MCMC の開発

内容: 数値実験による計算効率に関する検討

森 裕一 (岡山理科大学)

役割: 加速化 ALS アルゴリズムの開発

内容: 理論的評価と変数選択への適用および数値実験による性能評価

Geng Zhi (Peking University)

役割: 各研究課題に対する妥当性の評価と助言

4. 研究成果

研究目的に示した3つの課題それぞれについての成果を記す。

(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, ベクターイブ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* (2008) に掲載された。

(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* (2010) に掲載された。さらに, 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た。その成果をまとめた論文が, *Computational Statistics* (2012) に採択された。

(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* (2011) に発表した。

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 8 件)

- ① Kuroda, M., Mori, Y., Iizuka, M., and Sakakihara, M. (2012). Acceleration of convergence of the alternating least squares algorithm for nonlinear principal components analysis. *Principal Component Analysis* (P. Sanguansat Ed.), InTech Publications, 129-144. [査読有]
- ② Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M. (2011). Variable Selection in Principal Components Analysis of Qualitative Data Using the Accelerated ALS Algorithm. *Proceedings of the 7th conference of Asian Regional Section of IASC*, 83. [査読無]
- ③ Sakakihara, M. and Kuroda, M. (2011). King-Werner method for EM algorithm. *Proceedings of the 7th conference of Asian Regional Section of IASC*, 84. [査読無]
- ④ Kuroda, M., Iizuka, M., Mori, Y. and Sakakihara, M. (2011). Principal components based on a subset of qualitative variables and its accelerated computational algorithm. *Invited Paper Session 42, The 58th World Statistics Congress of the International Statistical Institute (ISI2011), Dublin, Ireland. (Invited paper)*. pp. 1-8. [査読有]
- ⑤ 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, pp. 143-153. [査読有]
- ⑥ 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* (L., Yves and S. Gilbert Eds.), Psysica-Verlag, Heidelberg, 1239-1246. [査読有]
- ⑦ 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]
- ⑧ 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]

- ⑨ 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]
- ⑩ Sakakihara, M. and Kuroda, M. (2008). Convergence of componentwise Aitken δ^2 acceleration of the EM algorithm. *COMPSTAT 2008: Proceedings in Computational Statistics* (P. Brito Ed.), Psysica-Verlag, Heidelberg, 597-903. [査読有]

[学会発表] (計 8 件)

- ① Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M. (2011). Variable selection in principal components analysis of qualitative data using the accelerated ALS algorithm. *7th Conference of Asian Regional Section of IASC*. 2012年12月16日. Academia Sinica (台湾).
- ② Sakakihara, M. and Kuroda, M. (2011). King-Werner method for EM algorithm. *7th Conference of Asian Regional Section of IASC*. 2012年12月16日. Academia Sinica (台湾).
- ③ Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M. (2011). Principal components based on a subset of qualitative variables and its accelerated computational algorithm. *Invited Paper Session 42, The 58th World Statistics Congress of the International Statistical Institute (ISI2011)*. 2012年8月22日. The Convention Centre Dublin (アイルランド).
- ④ 森 裕一, 飯塚誠也, 黒田正博, 榊原道夫 (2011). 質的データの主成分分析と計算の加速化. *日本計算機統計学会第25回大会*. 2012年5月8日. 函館市亀田福祉センター (北海道).
- ⑤ 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 (イギリス).
- ⑥ 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 (東京都).

- ⑦ 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(大阪府).
- ⑧ 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(京都府).
- ⑨ 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(神奈川県).

[その他]

ホームページ等

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

6. 研究組織

(1) 研究代表者

黒田 正博 (MASAHIRO KURODA)
岡山理科大学・総合情報学部・准教授
研究者番号: 90279042

(2) 研究分担者

森 裕一 (YUICHI MORI)
岡山理科大学・総合情報学部・教授
研究者番号: 80230085

榊原 道夫 (MICHIO SAKAKIHARA)
岡山理科大学・総合情報学部・教授
研究者番号: 70215614

中川 重和 (SHIGEKAZU NAKAGAWA)
倉敷芸術科学大学・産業科学技術学部・教授
研究者番号: 90248203