

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 14 日現在

機関番号：35302

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500353

研究課題名(和文)非計量データの変量解析における交互最小二乗法の加速に関する研究

研究課題名(英文)Acceleration of the alternating least squares algorithm for multivariate analysis of non-metric data

研究代表者

黒田 正博(Kuroda, Masahiro)

岡山理科大学・総合情報学部・教授

研究者番号：90279042

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：計量データを対象とした多変量解析法を非計量データへ適用するとき、最適尺度化にもとづく数量化をおこなう必要がある。本研究で対象としたのは、計量データの解析のための主成分分析および因子分析に加えて等質性分析(多重対応分析)である。計量データのための解析法において数量化計算を伴う場合、交互最小二乗法の適用が一般的である。しかし、この数値計算法は、最小二乗法に基づく反復法であるため収束スピードは線形である。そこで、交互最小二乗法の持つ収束特性を失うことなく収束スピードを改善する加速法を開発し、数値実験によりその推定性能を評価すると共に収束性と加速性の理論的考察をおこなう。

研究成果の概要(英文)：We developed acceleration methods of the alternating least squares (ALS) algorithms for nonlinear principal component analysis (NL-PCA) and factor analyses (NL-FA) of non-metric data. The vector epsilon accelerator is utilized for speeding up the convergence of the ALS algorithm. Numerical experiments demonstrated that the accelerated ALS algorithm provides less expensive computational cost than ordinary ALS algorithm in NL-PCA and NL-FA and greatly reduces the number of iterations and computational time of the ALS algorithm. These speeds of the acceleration algorithm are about 3 to 4 times faster than those of ordinary ALS algorithm.

研究分野：計算機統計学

キーワード：交互最小二乗法 収束の加速 ベクターイプシロン法 非計量主成分分析 非計量因子分析

1. 研究開始当初の背景

非計量データは順序尺度や名義尺度で測られたデータであり、主成分分析や因子分析といった計量データ(間隔・比尺度データ)を対象とした多変量解析法を直接適用することはできない。これらの手法を非計量データに用いるためには最適尺度化による数量化が必要になってくる。標準的な数量化法として、交互最小二乗法がある。この反復法は、適用する解析法で求めるべきパラメータと数量化によるデータの最適変換の計算を交互に繰り返し、これらの最小二乗解を求める。しかしながら、交互最小二乗法の収束スピードは最小二乗法と同じく1次であり、収束は遅い。したがって、個体および変数の数そして各変数のカテゴリー数が多い大規模データにおいて、この反復法により数値解を求めるためには、多大な反復回数と計算時間が必要になる。このため、交互最小二乗法の収束スピードを加速することは、大規模データの解析において重要な課題である。我々は、Kuroda et al. (2011)において、尺度混在データに対する主成分分析の交互最小二乗法の加速法を提案した(平成20年度~平成23年度:科学研究費(基盤(C))。そこで、本研究では、前述の研究を基に、非計量データの多変量解析法における交互最小二乗法の加速に関する研究に取り組んだ。

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) 今年度の研究では、非計量主成分分析に用いる交互最小二乗法の加速に焦点をあて、vector epsilon ($v\epsilon$)法による加速法の開発とその加速性能を数値実験により検証するとともに、実データによる変数選択問題に適用することで、この加速法の加速性能を検討した。

(2) 交互最小二乗法の加速法の優位性を示すために、収束性と加速性に関する理論研究をおこなった。我々が対象とした非計量主成分分析には、このパラメータの推定計算の過程で非線形方程式を解くステップがあるため、収束性を議論することが困難であった。そこで、各計算ステップで、非線形方程式による解法を用いない交互最小二乗法があることがわかったので、この反復法をベースにした研究をおこなった。

(3) 計量データと非計量データの混在したデータ(尺度混在データ)に対し、主成分分析を適用し、従来用いる交互最小二乗法と $v\epsilon$ 法による加速法を、計量データと非計量データの比率を変えての性能比較をおこなった。さらに、非計量データの因子分析にも、交互最小二乗法の $v\epsilon$ 法による加速化法の開発をおこなった。

3. 研究の方法

研究代表者の黒田および4名(榊原、森、足立、中川、飯塚)の研究分担者、そして研究目的を着実に達成するため、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: 交互最小二乗法の加速法の開発および数値実験による検証

榊原道夫(岡山理科大学)

役割: 交互最小二乗法の加速法の理論研究

内容: 交互最小二乗法の $v\epsilon$ 加速化法の収束性と加速性に関する理論研究

森 裕一(岡山理科大学)

役割: 交互最小二乗法の加速法の開発

内容: 非計量データに対する多変量解析法の交互最小二乗法の $v\epsilon$ 加速化法の開発

足立 浩平(大阪大学)

役割: 交互最小二乗法の加速法の開発

内容: 非計量データに対する多変量解析法の交互最小二乗法の $v\epsilon$ 加速化法の開発

中川重和(倉敷芸術科学大学)

役割: 交互最小二乗法の加速法の応用

内容: 非計量データの主成分分析における変数選択問題への $v\epsilon$ 加速化法の適用

飯塚 誠也(岡山大学)

役割: 交互最小二乗法の加速法の応用

内容: 非計量データの主成分分析における変数選択問題への $v\epsilon$ 加速化法の適用

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) 等質性分析で用いられる交互最小二乗法のアルゴリズム研究は、共同研究者である足立先生の著書およびこのアルゴリズムの開発者達の原著論文を調べることで完了した。 $v\epsilon$ 加速法の適用については、等質性分析のための交互最小二乗法のプログラムを完成することができなかったため、これについての成果を得ていない。そこで、共同研究者とともに至急にプログラムの完成を目指した。

非計量主成分分析の交互最小二乗法の収束性については、その証明を与えることができなかった。問題の1つは、交互最小二乗法の各反復で解く固有値問題の扱いである。この解析法に対する交互最小二乗法には、幾つ

かのバリエーションがあり、固有値問題を別の方法で回避しているアルゴリズムが存在することもわかった。そこで、これを非計量主成分分析の交互最小二乗法の基本アルゴリズムとして、この課題の解決を目指した。

大規模な乱数データによる変数選択問題において $v\varepsilon$ 加速法の性能を評価することができた。したがって、この課題は完了したと考えている。

(2) これまでの研究対象であった非計量主成分分析に対する交互最小二乗法とは別の定式化による交互最小二乗法のアルゴリズムと R 言語による実装をおこない動作を確認することができた。また、これらの方法を整理したサーベイ論文を書き、この反復法の収束性の証明に取り組むための準備は整った。そこで、交互最小二乗法の加速法の収束性と加速性に関する理論研究をおこなった。そのため、既存研究の成果をもとに交互最小二乗法の収束性に関する検討した。交互最小二乗法は、最小化する目的関数のパラメータを幾つかの群にまとめ、その群に属するパラメータごとに最小二乗推定を交互に適用することで目的関数を最小化する方法である。これまでの研究では、目的関数の最小化問題において交互最小二乗法が収束することを証明したものはあるが、目的関数のパラメータが収束することを証明した研究は存在しなかった。しかし、加速化法の収束理論に関する研究においては、パラメータの収束が保障され上での議論となるため、これを明らかにする必要がある。我々が対象とした非計量主成分分析には、このパラメータの推定計算の過程で非線形方程式を解くステップがあるため、収束性を議論することが困難であった。そこで、交互最小二乗法の中で非線形方程式を解かない数値解法に対し、 $v\varepsilon$ 加速法を適用し、それに対する収束性と加速性の理論研究をおこなった。

(3) 計量データと非計量データの混在したデータに主成分分析を適用する場面において、非計量データの最適スケールリングによる数量化が必要になる。これをおこなうための数値反復法として交互最小二乗法がある。このアルゴリズムは、非計量データを数量化し計量データとして推定するステップと、従来の主成分分析をおこなうステップからなる。このアルゴリズムの最適化基準として、最小二乗が用いられている。ただし、この基準をもとにした反復法は収束スピードが一次であり、非計量データの割合が計量データと比べ多い場合、収束までに多くの反復回数が必要になる。そこで、収束速度を改善するために提案したのが、ベクターイプシロンによる加速法であり、数値実験による検証では非計量データによる主成分分析において2倍以上で収束が速くなっていることを示した。さらに、計量データと非計量データの比率と、提案し

た加速法の収束速度の改善度合いの関係を数値実験により検証した。

また、非計量データに因子分析を適用した場合の交互最小二乗法の適用とベクターイプシロン法による加速法を提案し、小規模なデータにより、その加速の性能も検証をした。非計量データの因子分析において、交互最小二乗法の $v\varepsilon$ 法による加速が利用できることがわかった。ただし、因子分析のパラメータ推定法（最小二乗法および最尤法）の違いにより、加速法の加速の性能に違いがでることが数値実験により明らかになった。因子分析のパラメータ法の性質と加速法の関連について検証をすることが必要であると考えている。特に、不適解の場合についての処理は重要であるので、アルゴリズムの改良を検討する必要がある。これについては継続して研究を進めていく。

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 Kuroda M., Geng Z. and Sakakihara M. (2015). Improving the vector ε acceleration for the EM algorithm using a re-starting procedure. *Computational Statistics* 30, pp. 1051 - 1077. [査読有]
- ② Adachi K. (2015). A new algorithm for generalized least squares factor analysis with a majorization technique. *Open Journal of Statistics* 5, pp. 165-172. [査読有]
- ③ Mori Y., Kuroda M., Izuka M. and Sakakihara Y. (2014). Performance of acceleration of ALS algorithm in nonlinear PCA. *COMPSTAT 2014: Proceedings in Computational Statistics*, pp. 257-263. [査読有]
- ④ Adachi K. and Trendafilov N.T. (2014). Sparse orthogonal factor analysis: In *Advances in Latent Variables: Methods, Models and Applications*, Springer, pp. 227-239 [査読有]
- ⑤ 黒田正博 (2014). 相関・連関分析から因果分析へ. 数学セミナー増刊 統計学ガイドダンス, 日本評論社, pp. 120 - 125 [査読無]
- ⑥ Kuroda M., Mori Y., Izuka M. and Sakakihara M. (2013). Alternating least squares in nonlinear principal components. *WIREs Computational Statistics* 5, pp. 456 - 464. [査読有]
- ⑦ Kuroda M., Hashiguchi H., Nakagawa S. and Geng Z. (2013). MCMC using Markov bases for computing p-values in decomposable log-linear models. *Computational Sta-*

tistics 28, pp. 831- 850. [査読有]

- ⑧ Kuroda M., Mori Y., Izuka M. and Sakakihara M. (2013). Accelerating and re-starting the alternating least squares algorithm for non-linear principal components analysis. Proceedings of the 59th World Statistics Congress (ISI2013), pp. 5426-5431. [査読有]
- ⑨ Adachi K. (2013). Factor analysis with EM algorithm never gives improper solutions when sample covariance and initial parameter matrices are proper. Psychometrika 38, pp. 380-394. [査読有]
- ⑩ Adachi K. (2013). Generalized joint Procrustes analysis. Computational Statistics 28, pp. 2449-2464. [査読有]
- ⑪ 山本義郎, 飯塚誠也, 藤野友和 (2013). 統計データの視覚化 (R によるデータサイエンス 12). 共立出版. [査読無]
- ⑫ Kuroda M., Mori Y., Izuka M. and Sakakihara M. (2012). Acceleration of convergence of the alternating least squares algorithm for nonlinear principal components analysis. In Principal Component Analysis (Sanguansat, P. (Ed.)), InTech Publications, pp. 129-144. [査読有]
- ⑬ Kuroda M., Sakakihara M., Mori Y. and Izuka M. (2012). Two-stage acceleration for non-linear PCA. COMPSTAT 2012: Proceedings in Computational Statistics, pp. 461-471. [査読有]
- ⑭ Kohei A. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis. Journal of the Japanese Society of Computational Statistics 25, pp. 25-38. [査読有]
- [学会発表] (計 15 件)
- ① Kuroda M. (2015). Initial value selection of the EM algorithm for Gaussian mixture models. IASC-ARS, 2015 年 12 月 17 日～12 月 19 日, Singapore (Singapore).
- ② 黒田正博(2015). EM アルゴリズムの加速化と応用. 日本行動計量学会第 43 回大会, 2015 年 9 月 1 日～9 月 4 日, 首都大学東京(東京都八王子市).
- ③ 黒田正博(2015). 正規混合モデルの EM アルゴリズムの初期値選択と加速. 日本計算機統計学会第 29 回大会, 2015 年 5 月 14 日～15 日, 山梨県立図書館 (山梨県甲府市).
- ④ Kuroda M. (2014). An initial value selection

method of the EM algorithm for mixture models. Kyoto International Conference on Modern Statistics in the 21st Century, 2014 年 11 月 17 日～11 月 18 日, 京都国際会議場 (京都府京都市).

- ⑤ Mori Y., Kuroda M., Izuka M. and Sakakihara M. (2014). Performance of acceleration of ALS algorithm in nonlinear PCA. COMPSTAT2014 and IASC2014, Geneva (Switzerland).
- ⑥ 黒田正博(2014). 統計計算で用いる反復法の加速. 環瀬戸内応用数理研究部会第 18 回シンポジウム, 2014 年 12 月 5 日～12 月 7 日, 加計学園国際学术交流センター (岡山県倉敷市).
- ⑦ 黒田正博, 森裕一, 飯塚誠也, 榊原道夫 (2014). 非計量主成分分析における交互最小二乗法とその加速化. 2014 年度統計関連学会連合大会, 2014 年 9 月 13 日～9 月 16 日, 東京大学 (東京都文京区).
- ⑧ Kuroda M. (2013). Acceleration of the EM algorithm to mixture models. Ishigaki International Conference on Modern Statistics Theories, Practices, and Education in the 21st Century 2013, 2013 年 11 月 9 日～11 月 10 日, ANA インターコンチネンタル石垣リゾート(沖縄県石垣市).
- ⑨ Kuroda M., Mori Y., Izuka M. and Sakakihara M. (2013). Accelerating and re-starting the alternating least squares algorithm for non-linear principal components analysis. The 59th World Statistics Congress, 2013 年 8 月 25 日～8 月 30 日, Hong Kong (China).
- ⑩ Kuroda M. (2013). Acceleration of the EM algorithm using the vector epsilon accelerator and a re-starting procedure. The 59th World Statistics Congress, 2013 年 8 月 25 日～8 月 30 日, Hong Kong (China).
- ⑪ 黒田正博, 森裕一, 飯塚誠也, 榊原道夫 (2013). 非線形主成分分析の加速化: 変数選択への適用. 日本行動計量学会第 41 回大会, 2013 年 9 月 3 日～9 月 6 日, 東邦大学 (千葉県船橋市).
- ⑫ 黒田正博, 森裕一, 飯塚誠也, 榊原道夫 (2013). re-start を組み込んだ交互最小二乗法の加速法の非計量主成分分析への適用. 日本計算機統計学会大会第 27 回大会, 2013 年 5 月 6 日～5 月 7 日, 弘前大学 (青森県弘前市).
- ⑬ 黒田正博, 榊原道夫, 飯塚誠也, 森裕一 (2012). 非計量主成分分析の加速化—実データへ適用. 日本行動計量学会第 40

回大会, 2012年9月13日~9月16日, 新潟県立大学(新潟県新潟市).

- ⑭ 森裕一, 飯塚誠也, 黒田正博, 榑原道夫 (2012). Graves-Morris による交互最小二乗法の加速化. 日本計算機統計学会第26回大会, 2012年5月12日~5月13日, 香川県社会福祉総合センター(香川県高松市).
- ⑮ Kuroda M., Sakakihara M., Mori Y. and Iizuka M. (2012). Two-stage acceleration for non-linear PCA. COMPSTAT2012, 2012年8月27日~8月13日, Limassol (Cyprus).

[図書] (計 0 件)

[その他]

ホームページ等

<http://researchmap.jp/read0096675>

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・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

榑原 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

足立 浩平 (KOHEI ADACHI)

大阪大学・人間科学研究科・教授

研究者番号: 60299055

飯塚 誠也 (MASAYA IIZUKA)

岡山大学・アドミッションセンター・教授

研究者番号: 60322236