

平成26年度北海道大学情報基盤センター共同研究成果報告書

1. 研究領域番号 A3

2. 研究課題名 ビッグデータにおけるデータ行列の近似に関する総合的研究

3. 研究期間 平成26年 4月 1日 ～ 平成27年 3月31日

4. 研究代表者

氏 名	所属機関・部局名	職 名	備 考
宿久 洋	同志社大学・文化情報学部	教 授	

5. 研究分担者

氏 名	所属機関・部局名	職 名	備 考
足立 浩平	大阪大学・大学院人間科学研究科	教 授	
橋本 翔	大阪大学・大学院人文科学研究科	D3	
池本 大樹	大阪大学・大学院人文科学研究科	M2	
斉藤 光平	大阪大学・大学院人文科学研究科	M2	
宇野 光平	大阪大学・大学院人文科学研究科	M1	
寺田 吉壱	脳情報通信融合研究センター	研究員	
山本 倫生	京都大学・大学院医学研究科	特任助教	
谷岡 健資	同志社大学・大学院文化情報学研究科	D3	
南 弘征	北海道大学・情報基盤センター	准教授	

6. 共同研究の成果

下欄には、当該研究期間内に実施した共同研究の成果について、その具体的内容、意義、重要性等を、共同研究申請書に記載した「研究目的」と「研究計画・方法」に照らし、800字～1,000字で、できるだけ分かりやすく記載願います。文章の他に、研究成果を端的に表す図表を貼り付けても構いません。なお、研究成果の論文・学会発表等を行った実績（発表等の予定を含む。）があれば、あわせて記載して下さい。

近年、情報技術の発達に伴いデータが大規模複雑化し、データ解析による「予測」や「特徴把握」を行うことの需要が社会的・学術的にも高まってきている。そのような大規模複雑データに対して既存手法では、適用困難である場合や解析結果を得たとしてもその解釈が困難であるという問題が挙げられる。そこで、パラメータの推定が速く、モデルの解釈が容易である行列近似に基づくデータ解析手法の研究・開発は重要となる。そこで、研究代表者らは実際に有用で、かつ適用可能な行列近似に基づく新たな解析手法の開発を行った。

主な研究成果として、2つ挙げられる。1つめは、データ行列の対象数が多い場合、既存の次元縮約方法を適用しても結果の解釈が困難となる問題を克服するため、「クラスタリング法と次元縮約の同時分析法の開発」を行ったことである。各対象の特徴ではなく、各クラスターの特徴を解釈しやすい低次元空間を検知することが可能であり、かつクラスタリング結果の精度も良いことが利点として挙げられる。本研究成果では「量的データ」、「質的データ」、「関数データ」、「区間値」等に対する次元縮約クラスタリング法を提案し、いずれもその有効性を示すことができた。2つめは、変量数が膨大なデータ行列に対して、既存の次元縮約方法を適用しても結果の解釈が困難となる問題を克服するため「スパース推定に基づいた次元縮約法」を開発したことである。スパース推定によって各次元の解釈すべき項目数を減らすことができ、実用上適用可能な手法となっていることが利点として挙げられる。

成果の詳細は以下のとおりである。

（研究成果のつづき）

- [1] Yamamoto, M. and Hwang, H. (2014): A general formulation of cluster analysis with dimension reduction and subspace separation, *Behaviormetrika*, 41, 115-129.
- [2] Yamamoto, M. and Terada, Y. (2014): Functional factorial k-means analysis, *Computational Statistics & Data Analysis*, 79, 133-148.
- [3] Trendafilov, N. and Adachi, K. (2014): Sparse versus structure loadings, to appear in *Psychometrika*.
- [4] Mitsuhiro, M. and Yadohisa, H. (2014): Reduced k-means clustering with MCA in a low-dimensional space, to appear in *Computational Statistics*.
- [5] Adachi, K. and Trendafilov, N. (2014): Penalty-free sparse PCA, The 21st International Conference on Computational Statistics (COMPSTAT 2014), Geneva, Switzerland, p197-p203.
- [6] Abe, H., Tanioka, K. and Yadohisa, H. (2014): Clusterwise linear regression model for modal multi-valued data, *Proceedings of the International Conference on Mathematics, Statistics, and Financial Mathematics 2014*, p35-p40.
- [7] Adachi, K. and Trendafilov, N. (2014): Sparse SEM: computational identification of inter-factor relations in confirmatory factor analysis, *7th International Conference of the European Research Consortium for Informatics and Mathematics Working Group on Computational and Methodological Statistics 2014*, Pisa, Italy.
- [8] Uno, K. and Adachi, K. (2014): Fixed factor analysis with clustering observations, *7th International Conference of the European Research Consortium for Informatics and Mathematics Working Group on Computational and Methodological Statistics 2014*, Pisa, Italy.
- [9] Yamamoto, M. (2014): Functional multiple-set canonical correlation analysis for square integrable stochastic processes, The 21st International Conference on Computational Statistics (COMPSTAT 2014), Geneva, Switzerland.
- [10] Yamamoto, M. and Hayashi, K. (2014): Simultaneous analysis of clustering and dimension reduction for binary variables with application to biomedical data, *The 27th*

International Biometric Conference, Florence, Italy.

[11] Tanioka, K. and Yadohisa, H. (2014): K-mode clustering with dimensional reduction for categorical data. *European Conference on Data Analysis 2014*, Bremen, Germany.

[12] Takagi, I., Tanioka, K. and Yadohisa (2014): Constrained VPCA for interval-valued data, *2014 Workshop in Symbolic Data Analysis*, Taipei, Taiwan.

[13] Tsuchida, J. Yadohisa, H. (2014): Partial least squares logistic regression using F-measure, The 21st International Conference on Computational Statistics (COMPSTAT 2014), Geneva, Switzerland.

[14] Tsuchida, J. Yadohisa, H. (2014): Two-mode three-way asymmetric MDS using the log linear model, *7th International Conference of the European Research Consortium for Informatics and Mathematics Working Group on Computational and Methodological Statistics 2014*, Pisa, Italy.

[15] Abe, H., Tanioka, K., and Yadohisa, H. (2014): Reduced k -means for multivalued quantitative symbolic variables, *7th International Conference of the European Research Consortium for Informatics and Mathematics Working Group on Computational and Methodological Statistics 2014*, Pisa, Italy.

[16] Arishige, B. and Yadohisa, H. (2014): Clusterwise sparse principal component regression, *7th International Conference of the European Research Consortium for Informatics and Mathematics Working Group on Computational and Methodological Statistics 2014*, Pisa, Italy.