別　紙　３

平成２７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　　行列分解に基づく大規模複雑データ解析法に関する研究

３．研 究 期 間　　平成２７年　４月　１日　～　平成２８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
水田　正弘	北海道大学・情報基盤センター	教授	
足立　浩平	大阪大学・人間科学	教授	
山本　倫生	京都大学・医学研究科	特定助教	
寺田　吉壱	脳情報通信研究機構・脳情報通信融合研究センター	研究員	
阿部　寛康	同志社大学・大学院文化情報学研究科	D2	
池本　大樹	大阪大学・大学院人間科学研究科	D1	
中村　裕子	大阪大学・大学院人間科学研究科	M2	
宇野　光平	大阪大学・大学院人間科学研究科	M2	

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てきており，既存の解析手法では目的を達成できない場合がある．加えて，様々な解析手法が提案されており，そのような手法間の関係性を明らかにし，体系化することが望まれている．これまで，研究代表者らは主として行列近似問題として定式化した新たな多変量データ解析法を提案し，その有効性や既存手法との関連を明らかにすることに取り組んできた．今回は特に大規模複雑データ特有の問題を考慮した行列分解型多変量データ解析法の提案とその理論的側面の追求を目的として研究に取り組んだ.

主な研究成果として，2 つ挙げられる．1 つ目は，分析目的に沿って形状が異なるような様々なタイプの大規模複雑データに対する行列分解型多変量解析法を開発したことである．例えば，分類を伴う次元縮約法においては，全ての変数が順序尺度である多変量データや，非対称非類似度データ，キューブ状に与えられる多元データなどに対して適切な手法を開発した．これらは既存の手法では解釈可能な結果を抽出できない状況に適応した手法である．またほかにも，2 つの多変量データのうちひとつからクラスターを抽出し，そのクラスターをラベルとみなしてもう片方のデータで判別する解析法や，全てが非負値で与えられるデータに適した行列分解型解析法，関数データに対する正準相関分析など，目的や状況によって特殊な形状・性質をもつデータに対する新しい解析法を開発した．研究成果の 2 つ目は，行列分解型多変量データ解析法に関する理論的側面を明らかにしたことである. 寺田はクラスタリング法と次元縮約法の同時分析法のひとつである Factorial K-means 法によって得られる推定値が概収束するための条件を導いた.

（研究成果のつづき）

成果の詳細は以下のとおりである.

[1] Terada, Y. (2015): Strong consistency of factorial k-means clustering, *Annals of the Institute of Statistical Mathematics*, 67 (2), 335-357.

[2] Yamamoto, M. and Hayashi, K. (2015): Clustering of multivariate binary data with dimension reduction via L1-regularized likelihood maximization, *Pattern Recognition*, 48, 3959-3968.

[3] Hirose, K. and Yamamoto, M. (2015): Sparse estimation via nonconcave penalized likelihood in a factor analysis model, *Statistics and Computing*, 25, 863-875.

[9] Mitsuhiro, M. and Yadohisa, H. (2015): Reduced k-means clustering with MCA in a low-dimensional space, *Computational Statistics*, 30 (2), 463-475.

[10] Tanioka, K. and Yadohisa, H. (2016): Discriminant coordinates for asymmetric dissimilarity data based on radius model, *Behaviormetrika*, 42(2), p1-17.

[4] Yamamoto, M. and Terada, Y. [invited] (2015): Canonical correlation analysis for multivariate functional data, *8th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics)*, Senate House, University of London, UK.

[5] Yamamoto, M., Kawaguchi, A., and Hwang, H. (2015): Outcome-guided clustering using supervised dimension reduction approach, *2015 International Workshop for JSCS 30th Anniversary*, Okinawa Institute of Science and Technology Graduate University, Japan.

[6] Yamamoto, M. and Kawaguchi, A. (2015): A component-based approach to find outcome-related clusters, *The 80th Annual meeting of the Psychometric Society (IMPS 2015)*, Beijing Normal University, Beijing, China.

[7] Yamamoto, M. (2015): A simultaneous analysis of dimension reduction and clustering with correlated error variables, *The 2015 conference of the International Federation of Classification Societies (IFCS 2015)*, Bologna, Italy.

[8] Adachi, K. (2015): A New Algorithm for Generalized Least Squares Factor Analysis with a Majorization Technique, *Open Journal of Statistics*, 5(3), 165-172.

[11] Ikuta, H., Tanioka, K. and Yadohisa, H. (2015): Simultaneous analysis of multi-label classification and dimensionality reduction with clustering labels, *60th World Statistics Congress*, Rio de Janeiro, Brazil.
[12] Tsuchida, J. and Yadohisa, H. (2015): Bayesian Asymmetric Multidimensional Scaling for Two-mode Three-way Count Data by Using Log-linear Model, *The 9th Conference of the Asian Regional Section (ARS) of the International Association for Statistical Computing (IASC)*, Shingapore.
[13] Tanioka, K. and Yadohisa, H. (2015): Constrained asymmetric MDS based on raidus model, The 9th Conference of the Asian Regional Section (ARS) of the International Association for Statistical Computing (IASC), Shingapore.
[14] Tsuchida, J. and Yadohisa, H. (2015): Canonical correlation analysis for three-mode three-way data, The 2015 conference of the International Federation of Classification Societies, p253, Bologna, Italy.
[15] Tsuchida, J. and Yadohisa, H. (2015): Bayesian unfolding for count data by using log-linear model, Statistical Computing Asia, p13, Taipei, Taiwan.
[16] Tanioka, K. and Yadohisa, H. (2015): Clustering for categorical data based on non-negative subspace, The 24th South Taiwan Statistics Conference and 2015 Chinse Institute of Probability and Statistics Annual Meeting, p95, Changhua, Taiwan.
[17] Abe, H. and Yadohisa, H. (2015): Automatic Relevance Determination in NMF based on a Zero-Inflated Compound Poisson Model, *2015 International Workshop for JSCS 30th Anniversary in Okinawa*, Okinawa Institute of Science and Technology Graduate University.
[18] Hasegawa, K. and Yadohisa, H. (2015): Multidimensional scaling for one-mode three-way symbolic dissimilarity data, The 9th Conference of the Asian Regional Section (ARS) of the International Association for Statistical Computing (IASC), Shingapore.