

## 平成25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複雑類似性データの解析法に関する総合的研究
3. 研究期間 平成25年 4月 1日 ~ 平成26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	
波多野 賢治	同志社大学・文化情報学部	准教授	
深川 大路	同志社大学・文化情報学部	助教	
光廣 正基	同志社大学・文化情報学研究科	M2	
谷岡 健資	同志社大学・文化情報学研究科	D2	
有重 文平	同志社大学・文化情報学研究科	M1	
山本 康平	中央大学・理工学研究科	M2	
秋元 良友	中央大学・理工学研究科	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、計算機やインターネット等の情報機器の発展・普及によって、大量のデータが高頻度に蓄積されるようになり、購買履歴データやアクセスログデータのようなビッグデータの活用が注目視されている。ビッグデータをリアルタイムに解析する技術が拡大し始め、ストリーミング解析等が有効であると考えられている。しかし、あらゆる分野で扱われているデータが大規模かつ複雑なものになっているため、従来のデータ解析手法では適用困難であり、大規模複雑なデータに対応する解析手法が必要とされている。なかでも、さまざまな現象を扱うような2者関係を示すデータ（類似性データ）が多種多様に存在しているため、それらを分析する需要は高い。

本研究では、通常の変量データと異なる類似性データについて、内在する複雑な構造を考慮した新たな解析手法及びその解析手法に関する評価方法の提案・検証を目的として研究に取り組んだ。類似性データは対象×対象の形の正方行列で与えられ、1) 対象間の関係性が「対称」や「非対称」、2) データ形式として他者関係や他者間と他の要因を考慮した「多相多元データ」、3) データ値が区間値や分布値等で記述される「シンボリックデータ」、4) ビッグデータ特有の「スパースデータ」や「タイデータ」の存在など考慮すべき点は多い。

本研究の成果として、通常的手法では解析困難な大規模複雑類似性データへの適応可能な新手法の提案を行った。成果の詳細は以下のとおりである。

(研究成果のつづき)

- [1] Takagishi, M., Hirotsuru, K., Kusaka, T., Mitsuhiro, M. and Yadohisa, H. (2013): Educational feature extraction across nations using UNdata, Proceedings of Joint meeting of the IASC satellite conference for the 59th ISI WSC and the 8th conference of the asian regional section of the IASC, Seoul, Korea, p429-p431, (IASC Data Analysis Competition Award).
- [2] Tanioka, K. and Yadohisa, H. (2013): Three-Mode Hierarchical Subspace Clustering with Noise Variables and Occasions, German-Japanese Interchange of Data Analysis Results (W, Gaul et al. eds), Springer, Heidelberg, p91-p99.
- [3] Mitsuhiro, M. and Yadohisa, H. (2013): Multiple Correspondence Analysis for Mixed Measurement Level Data, European Conference on Data Analysis 2013, Luxembourg, Luxembourg.
- [4] Umei, T. and Yadohisa, H. (2013): Non-hierarchical clustering algorithm for mixed numerical and categorical three-way three-mode data, conference of the International Federation of Classification Societies, Tilburg, the Netherlands.
- [5] Tanioka, K. and Yadohisa, H. (2013): Ultrametric tree representation for three-way three-mode data with weights of variables and occasions, conference of the International Federation of Classification Societies, Tilburg, the Netherlands.
- [6] Yamashita, Y. and Yadohisa, H. (2013): MDS for series data by using candlestick valued dissimilarity measure, the 78th annual meeting of the psychometric society, Arnhem, the Netherlands.
- [7] Mitsuhiro, M. and Yadohisa, H. (2013): Simultaneous Fuzzy Clustering with Multiple Correspondence Analysis, the 59th World Statistics Congress of the ISI, Hong Kong, China.
- [8] 谷岡健資, 宿久洋 (2013): 3相3元データに対する変量および条件の重みを考慮した階層的クラスタリング法について, 日本分類学会第31回大会, (於 中央大学).
- [9] 梅井隆弘, 宿久洋 (2013): 量質混在3相3元データに対する非階層クラスタリング法. 日本計算機統計学会 第27回大会講演論文集 p105-108, (於 弘前大学).
- [10] 光廣正基, 宿久洋 (2013): 対象の分類を伴う多重対応分析法. 日本計算機統計学会 第27回大会講演論文集 p51-54, 学生研究発表賞, (於 弘前大学).
- [11] 山下陽司, 宿久洋 (2013): 系列データに対するトレンドとばらつきを考慮した多次元尺度構成法について. 日本計算機統計学会 第27回大会講演論文集 p35-38, (於 弘前大学).
- [12] 光廣正基, 宿久洋 (2013): 大学生協食堂のPOSデータ解析 —ミールカード利用者の食の実態把握をめざして—. 2013 PCカンファレンス, 最優秀論文賞, (於 東京大学).
- [13] 山下陽司, 宿久洋 (2013): 経時的に得られた非対称非類似度データに対するトレンドとばらつきを考慮した多次元尺度構成法について. 日本行動計量学会, (於 東邦大学).
- [14] 谷岡健資, 宿久洋 (2013): 3元データ分析のための additive tree 推定と視覚化について. 2013年度統計関連学会連合大会, (於 大阪大学).
- [15] 有重文平, 宿久洋 (2013): 多変量因子回帰分析法の提案. 日本計算機統計学会 第27回シンポジウム講演論文集 p23-26, 学生研究発表賞受賞, (於 熊本市民会館).
- [16] 浅野祐介, 谷岡健資, 宿久洋 (2013): カテゴリカル3相3元データの分析法について. 日本計算機統計学会 第27回シンポジウム講演論文集 p31-34, (於 熊本市民会館).