

平成22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 大規模データ科学
2. 研究課題名 大規模複雑データ解析に関する研究
ーシンボリックデータ解析法によるアプローチー
3. 研究期間 平成22年10月1日 ～ 平成23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	
南 弘征	北海道大学・情報基盤センター	准教授	
酒折 文武	中央大学・理工学部	准教授	
高木 育史	同志社大学・大学院文化情報学研究科	M1	
Park Heewon	中央大学・大学院理工学研究科	D1	
寺田 吉吉	同志社大学・大学院文化情報学研究科	M2	

6. 共同研究の成果

シンボリックデータとは、データの表現を一般化するための概念であり、従来のデータでは、データ点が実数空間の一点で表現されるのに対して、シンボリックデータではデータ点の値としてシンボリックオブジェクトといわれる区間値、多値、modal 値などを仮定している。シンボリックデータ解析は、シンボリックデータに対する解析手法であり、従来のデータ解析手法をシンボリックデータに対するものへと拡張するものとなっている。

本研究では、1) シンボリックデータの理論的特徴付け、2) 既存のシンボリックデータ及び大規模データに対するデータ解析法の総合的な調査、3) シンボリックデータ及び大規模データに対する新しい解析法の提案を通して、シンボリックデータ解析を中心に大規模データに対する解析手法の研究に取り組んだ。

特に本年度は、これらのテーマについて、2回の研究討論会を行い、集中的に最新の研究成果の相互レビューを実施した。主な研究成果は以下のとおりである。

[1] Terada, Y. and Yadohisa, H. (2010). Non-Hierarchical Clustering for Distribution-Valued Data. to appear in COMPSTAT 2010: Proceedings in Computational Statistics, Psysica-Verlag, Heidelberg, p1653-p1660.

- [2] Terada, Y. and Yadohisa, H. (2010). Modal Interval-Valued Dissimilarity Between Histogram-Valued Data. 3rd German-Japanese Workshop "ADVANCES IN DATA ANALYSIS AND RELATED NEW TECHNIQUES AND APPLICATIONS", Karlsruhe, German.
- [3] Terada, Y. and Yadohisa, H. (2010). Kernel Methods for Analyzing Symbolic Data. GfKI 2010, Karlsruhe, German.
- [4] Tanioka, K. and Yadohisa, H. (2010), Effect of Data Standardization on the Result of k -Means Clustering, GfKI 2010, Karlsruhe, German.
- [5] Saito, Y. and Yadohisa, H. (2010). Visualization of Asymmetric Clustering Result with Digraph and Dendrogram, GfKI 2010, Karlsruhe, German.
- [6] Terada, Y. and Yadohisa, H. (2010), Principle Component Analysis for Histogram-valued Data, The 10th China-Japan Symposium on Statistics Collected Papers, 254-257.
- [7] 寺田吉壱, 宿久洋. (2010). 分布値データに対するクラスタリング法の提案. 日本計算機統計学会第24回大会報告集, p1-p4.
- [8] 齊藤裕一, 宿久洋. (2010). 非対称階層的重複クラスタリング法の提案. 日本行動計量学会第38回大会発表論文抄録集, p88-p89.
- [9] 齊藤裕一, 寺田吉壱, 宿久洋. (2010). 対象の密集度を考慮した重複クラスタリング法の提案. 日本分類学会 28回研究報告会 予稿集, p15-18.
- [10] 寺田吉壱, 宿久洋. (2010), 区間値非類似性データに対する超球モデル MDS について. 日本分類学会 第27回研究報告会予稿集, p35-p38.
- [11] 寺田吉壱, 宿久洋. (2010), ヒストグラム値データに対する主成分分析法について. 2010年度統計関連学会連合大会講演報告集, , p282.
- [12] 阿部寛康, 池田健人, 波多野賢治, 宿久洋. (2010), 内容ベースフィルタリングを用いた動画検索サイトの利用履歴分析. 日本分類学会 第27回研究報告会予稿集, p31-p34.