

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 大規模データに対するシンボリックデータ解析法に関する研究
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
清水 信夫	統計数理研究所・データ科学研究系	助教	

5. 研究分担者

氏名	所属機関・部局名	職名	備考
片山 琴絵	東京大学・医科学研究所	特任研究員	
中野 純司	統計数理研究所・モデリング研究系	教授	
南 弘征	北海道大学・情報基盤センター	准教授	
水田 正弘	北海道大学・情報基盤センター	教授	
馬場 康維	統計数理研究所・統計思考院	特命教授	

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。

大規模かつ複雑なデータに対する解析方法の一つとして、シンボリックデータ解析法が研究されている。シンボリックデータ解析法は、解析対象を集合や抽象的な概念ととらえるという特徴があり、解析対象の記述方法として、区間値、分布値、尺度混合値など多様なものが提案されている。これらは、大規模データの簡潔な記述にも適していると考えられる。また、具体的な解析手法として、主成分分析をはじめとする次元縮小法、回帰分析などの各種推測手法、クラスター分析をはじめとする各種分類手法などがあり、各記述方法に対する解析手法について多くの研究が進められている。

本研究課題においては、シンボリックデータ解析法において大規模データの記述方法および解析手法の拡張を行った。特に、大規模データに含まれる様々な異なる小集団の特徴を生かしつつ解析における計算時間の削減にも寄与する記述方法を検討した。また、そのような記述方法を用いた解析手法のうち、クラスター分析の拡張を中心に検討し、これと関連がある多次元尺度構成法などの各種手法の拡張についても考察した。

成果として、研究代表者はシンボリックデータの一類型と考えられ得る集約的シンボリックデータ(ASD)という概念を提唱し、それを用いて大規模データに含まれる様々な異なる小集団の特徴を簡潔かつ十分に記述しつつクラスター分析を行う手法について提案した。また、それらについての研究報告を国内外における学会およびワークショップにおいて行い、他の研究者との討論を通じて可能性および課題について考察し理解を深めた。

(研究成果のつづき)

さらに、2013年12月16日に北海道大学情報基盤センターにて研究集会を行い、そこで研究代表者および分担者等を交えた様々な発表および総合的な討論を通じて先行研究事例および新たな研究の可能性について考察を行った。

今後はこれらの成果を元に、大規模かつ複雑なデータの記述法やクラスター分析手法についてさらなる改良および拡張を行い、シンボリックデータ解析法のさらなる発展を目指したいと考えている。