

平成24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 複雑な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24年4月23日 ~ 平成25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
東海林 智也	函館工業高等専門学校 情報工学科	准教授	
Myung-Hoe Huh	高麗大学 統計学部	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	

6. 共同研究の成果

大規模で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ある。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にとらえることに特徴があり、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ある。また、具体的な解析手法としては、主成分分析をはじめとする次元縮小法、推測のための回帰分析、分類のためのクラスター分析などがあり、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、分布値により記述されたシンボリックデータを中心に、クラスター分析の拡張を行うことを目的とする。特に、各解析対象が分布で記述された場合と、解析対象間の非類似度が分布で記述された場合について検討する。また、関連手法として、分布値データに対する多次元尺度構成法についても考察を行う。

成果として、研究代表者は分布値により記述された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、その上で ASD のクラスタリング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。

さらに、2012年7月10日および11日に、大韓民国より高麗大学の Myung-Hoe Huh 教授をはじめとする研究者および学生を北海道大学に招き、北大情報基盤センターと高麗大学との合同の研究集会を実施した。そこで研究代表者および分担者も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、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。