

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 A 3 大規模データ科学
2. 研究課題名 先端的クラスター分析に関する研究
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
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
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	
宿久 洋	同志社大学・文化情報学部	教授	

6. 共同研究の成果

クラスター分析とは、データを一定の類似性に基づいて少数の集団（クラスター）に分類するための解析手法であり、データ間の階層関係を見出す手法である階層的クラスター分析と、データ間の階層関係を仮定しない形での分類結果を見出す手法である非階層的クラスター分析に大別される。

従来のクラスター分析では、それぞれの分析手法において実数空間上の一点として表されるデータ点を扱う手法が研究されてきたが、近年ではインターネットや計算機の発展、普及によって、様々な分野で扱われるデータが大規模化かつ複雑化している。中でも、単一値としてでは表せない、区間値、多値、モーダル値などで表されるデータが多く出現するようになっており、そのような様々なデータに対するクラスター分析の需要も高まっている。

これらのデータの表現を一般化するための概念としてはシンボリックデータが提案されており、既存の様々なデータ解析手法をシンボリックデータに適用する研究が進められているが、本研究では、1) クラスター分析手法の理論的特徴付け、2) 従来型のデータに対するクラスター分析手法の総合的な調査、3) 海外における先行研究事例の学習、4) 様々なシンボリックデータに対する新たなクラスター分析手法の提案、を行うことによりクラスター分析においてもシンボリックデータに対して拡張するための研究を行い、クラスター分析法をさらに発展させることを目的とする。

成果として、研究代表者はシンボリックデータ解析の枠組に沿って、シンボリックデータの一様とみなし得る区間値関数データに対してクラスター分析を行う手法を提案し、今後の国際会議等で報告を予定している。また、シンボリックデータに対するクラスター分析の研究に関して多くの業績が有るポルト大学（ポルトガル）の Paula Brito 准教授を2011年2月上旬に北海道大学に招き、基調講演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した。

今後はこれらの成果を元に、クラスター分析手法の更なる改良や様々なデータへの拡張を行い、クラスター分析法のさらなる発展を目指したいと考えている。