

Title	教師なし分類学習に関する計量心理学的研究
Author(s)	西田, 豊
Citation	大阪大学, 2013, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/27494
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	にしだ豊
博士の専攻分野の名称	博士(人間科学)
学位記番号	第26072号
学位授与年月日	平成25年3月25日
学位授与の要件	学位規則第4条第1項該当 人間科学研究科人間科学専攻
学位論文名	教師なし分類学習に関する計量心理学的研究
論文審査委員	(主査) 教授 足立 浩平 (副査) 教授 狩野 裕 准教授 篠原 一光

論文内容の要旨

第1章 序章

分類とは物や現象を種類や性質などの基準により、似たもの同士をまとめ、似ていないものを区別することである。分類の目的は、多数の対象を把握するために、少数の群に分けて整理することである。分類のための基準を構成することを分類学習といい、外的基準の有無によって、教師あり学習と教師なし学習の2種類に分けられる。分類に関する研究は、生物学をはじめ、統計学、心理学など多くの分野で研究がなされてきた。本論文では、データ構造の探索を目的とする教師なし分類学習、すなわちクラスタリングを計量心理学の立場から扱う。クラスタリングの代表的手法に、K平均法、階層的クラスタリング、混合分布を用いたクラスタリングがある。

探索的に用いられるクラスタリング法においては、パラメータがファジィな連続値で与えられるファジィクラスタリングを用いたほうが、柔軟な解釈を行うことができると考えられる。本論文の目的は、ファジィ理論を用いて、結果の解釈を容易にし、より情報を引き出すことができる新たなクラスタリング法を開発することである。

第2章 数学的準備

はじめに、行列・ベクトルの表記法、基本演算、基本的定義を整理し、逆行列、固有値分解、特異値分解などについてまとめた。次に、パラメータ推定に必要なラグランジュ乗数法と正則化法を概説した。最後に、クラスタリングの分類結果の評価指標となるAdjusted Rand Indexを紹介した。

第3章 ファジィクラスタリング

K平均法では、個体がクラスターに所属するか否かは、0か1の値で与えられ、明確に分類される。しかしながら、実際のデータでは、クラスターへの所属が明確でない場合もある。ファジィc平均法は、ファジィ理論を用いて、個体がクラスターに所属する程度を0以上1以下の数値で表現できるよう、K平均法の制約を緩和したものである。本章では、べき乗型ファジィc平均法、エントロピー正則化ファジィc平均法、および、クラスターサイズ調整エントロピー正則化ファジィc平均法を概観した。

第4章 K平均法における変数選択法の開発

我々が手にするデータには、クラスター構造に関与する本質的な変数と、クラスター構造には関係のない変数が混在していることがある。このようなデータにおいては、クラスター構造に関与する変数のみを用いてクラ

スタリングを行ったほうが、良い結果を得られる場合がある。本章では正則化法を用いたK平均法における変数選択法を開発した。

まず、K平均法においては潜在的に固定化されている、変数に対する係数パラメータを顕在化した定式化を行った。そして、ファジィ理論を用いることにより、変数に対する係数パラメータを推定可能なK平均アルゴリズムを開発した。数値実験の結果より、提案手法は通常のK平均よりも高い精度でクラスタリング可能であり、有用な手法であることが示された。データ解析の結果より、提案手法は変数に対する係数パラメータを解釈することにより、変数選択を可能にし、クラスターの特徴を容易に理解できることが示された。さらに、本章で提案された変数選択をとまなうK平均法に、クラスターサイズ調整パラメータを導入した手法を開発した。数値実験の結果、通常のK平均法や変数選択をとまなうK平均法よりも良い成績を示すことが明らかとなった。

第5章 データ行列の最小二乗近似によるファジィクラスタリング法の開発

従来のファジィc平均法では、K平均法の目的関数にファジィパラメータの導入や、正則化項を加えることによって、目的関数を変形させ、メンバーシップパラメータのファジィ化を達成していた。本章ではK平均法の目的関数を変形させることなく、直接ファジィメンバーシップ値を推定する手法を開発した。提案手法は、従来のファジィc平均法と異なり、事前に設定しなければならないチューニングパラメータが存在しない。

まず、行列表現によるK平均法の定式化を用いることによって、従来行われてきたファジィc平均法とは異なるファジィクラスタリング法の目的関数を提案した。そして、メンバーシップパラメータをリパラメータライズすることにより、ファジィメンバーシップを推定する交互最適化アルゴリズムを開発した。数値実験の結果より、提案手法は、クラスタリング手法として有用であることが示された。データ解析の結果より、提案手法は主成分分析と類似した結果を出力することが示された。提案手法は主成分分析と目的関数が一致し、制約条件のみが異なるため、ファジィクラスタリングの特徴と、主成分分析の特徴をあわせ持った中間的手法といえる。

第6章 次元縮約をとまなうファジィK平均法の開発

ファジィクラスタリングには、クリスプクラスタリングよりも、局所解が低減される、妥当な解釈が可能、など様々なメリットがある。本章では、次元縮約とクラスタリングを単一の目的関数の最小化によって達成する次元縮約K平均法のファジィ拡張を行った。

まず、エントロピー正則化による次元縮約ファジィK平均法の目的関数を提案した。そして、パラメータを推定する反復アルゴリズムを開発した。数値実験の結果により、提案手法はノイズ変数が多い条件でも高い精度で分類可能であることが示された。データ解析の結果より、クラスターが明確に分離していないとき、ファジィなメンバーシップ値が解釈に有用であることが示された。さらに、クラスターサイズを調整する機能を有した次元縮約ファジィK平均法の目的関数を提案し、パラメータの推定アルゴリズムを開発した。データ解析によりその有用性が示された。

第7章 総合考察

クラスタリング手法は、外的変数を用いない教師なし学習による、対象のグループ分けであり、探索的に用いられる。ファジィ化の対象となるパラメータと、ファジィ化の方法という2つの観点から、ファジィ理論を用いたクラスタリング手法を分類し、解析手法の位置づけを行った。本論文で提案した手法は、連続的なパラメータ値を得ることによって柔軟な解釈が可能になっており、探索的データ解析において有用である。

数理モデルを扱う心理学関連分野として、心理データの解析手法の開発を目的とする計量心理学と、数理モデルによる心理現象の解明を目的とする数理心理学があるが、両者の厳密な線引きは難しい。そこで、教師なし分類学習の手法をデータ解析手法ではなく、計算論的認知モデルとして扱うことの可能性を考察した。人間が教師なし学習状況において概念を形成することは、事例間の類似性から最適な分割を発見することすることに他ならず、クラスタリングといえる。第4章や第6章で提案した手法は、人間が行なっているクラスタリングの特徴として仮定される、概念の曖昧さや特徴次元の選択・縮約を備えている。したがって、これらの手法は、人間の教師なし概念形成モデルとしても扱うことができると考えられる。

クラスタリング手法には、解決されていない問題もあるが、そのような問題に対処する新たな方法や、解釈を容易にし、より精度の高い分類を可能にする方法が多く提案されている。たとえば、非線形な分類境界を得られるカーネル法を用いた手法、部分的に得られている外部情報を積極的に用いる制約付きクラスタリングや半教師つきクラスタリング手法、クラスタリングと他の解析手法を組み合わせて同時解析を行う手法、などがある。本論文で提案した手法と上記のような方法を組み合わせて、新たなクラスタリング法を開発することで、より良い解釈を行う事ができると考えられる。最後に、 K 平均法において、いまだ未解決の大きな問題は、定番と呼ばれるクラスター数の決定法がないことである。理論的に整備されたクラスター数決定法の開発が望まれる。

論文審査の結果の要旨

個体の分類基準が与えられない状況で、類似した個体どうしを同じ群に、相違する個体どうしを異なる群に分類する作業は、認知科学や工学系の分野では教師なし分類学習と呼ばれ、多変量統計解析の分野ではクラスター分析と呼ばれる。本論文は、クラスター分析の諸方法の中でも、ポピュラーな非階層的方法である K -平均クラスタリングとその拡張であるファジィ・クラスタリングに焦点をあて、心理学データへの適用を考慮しながら、新たなクラスター分析の方法の研究開発を記したものである。開発された方法は、[1] オリジナルの変数群から分類に役立つ変数群だけを取捨選択する方法、[2] 新たなファジィ・クラスタリングの手法、および、[3] 個体のファジィな分類と同時に縮約次元の空間表現を行う手法であり、これら[1]、[2]、[3]に関する研究成果は、以下のように要約される。

[1] K -平均クラスタリングにおいて、分類に役立つ変数を取捨選択するために、分類に対する変数の寄与度を表すウェイトを未知パラメータとして推定するアルゴリズムを開発している。この手法によって、分類に役立つ変数のウェイトには高い値が推定され、その結果、高精度の個体の分類が実現されることを確認している。さらに、推定されたウェイトは、分類結果のクラスターの特徴の解釈にも役立つことが例証されている。また、開発手法のアルゴリズムでは、クラスターのサイズを調整するパラメータが使われるが、このことが、クラスター間でサイズが均等化されやすいという K -平均クラスタリングのバイアスを低減させるという効用をもつことも示されている。

[2] ある群に個体が所属するか否かの二値分類ではなく、各群への個体の所属の程度を0以上1以下の連続量として推定する新たなクラスター分析法を考案している。同様に連続量で群への所属度を表すパラメータを求める既存手法にファジィ c -平均法があるが、この方法の欠点は、事前にユーザーが主観的にその値を決めなければならないチューニングパラメータを含む点である。一方、そうした必要がない点が、本研究で開発された手法の長所といえる。本手法は、主成分分析に制約を加えた特殊ケース、および、 K -平均クラスタリングの一般化と位置づけられ、主成分分析とクラスター分析の数理的關係づけに新たな洞察を与える可能性を秘めている点でも意義深い。

[3] 個体の群への分類と、個体を位置づけた空間座標を求めるという、互いに異なる二つの目的を単一の計算で同時に達成する多変量解析法は既に考案されているが、それらは、個体がある群に所属するか否かの二値分類を行うものであった。本研究では、各群への個体の所属度を0以上1以下の連続量で表すと同時に空間座標を求める解析法が開発された。そのアルゴリズムの特徴は、個体と群の間のエントロピーを大きくするためにペナルティ関数を利用する点にある。以上の方法は、複数群の境界に位置する個体の自然な分類と、個体どうしの関係の視覚的な把握を可能にさせるものである。

行動科学をはじめ、いかなる分野においても、対象の分類は研究の初期段階で必須の作業となるため、教師なし分類学習・クラスター分析のニーズは絶えることはない。こうした統計解析の新たな手法の研究開発が精緻に論じられ、開発手法の計量心理学的位置づけ、さらには、認知科学との関わりの意義深い考究もなされた本論文は、博士（人間科学）の学位授与に十分に値するものと判定された。