氏名	中　村　永　友
学位（専攻分野）	博士（学術）
学位記番号	総研大甲第１０９号
学位授与の日付	平成７年３月２３日
学位授与の要件	数物科学研究科　統計科学専攻 学位規則第４条第１項該当
学位論文題目	クラスター化法の統計的評価とその応用
論文審査委員	主　査　教　授　駒　澤　　勉 教　授　大　隅　　昇 教　授　村　上　征　勝 助教授　馬　場　康　維 教　授　柳　井　晴　夫 （大学入試センター） 教　授　小　西　貞　則（九州大学）

論文内容の要旨

本論文は，多変量特性データの分類法であるクラスター化法（とくに組み合わせ的階層分類法）ならびに多変量混合分布モデルに基づく分類手法に注目して，これら両者の分類法としての特性を数理的に考察すると共に，両者の利点を活かした新たな分類方式を提案している．

論文は第Ⅰ部から第Ⅳ部，併せて10章から構成される．

第Ⅰ部（１章）の総論では，この研究に至った経緯，問題の背景について述べている．とくに，階層的分類法および多変量混合分布モデルの従来の研究における問題点と本論文で扱う研究の範囲，その意義等を議論している．

第Ⅱ部（２章～５章）では，階層的分類法のクラスター化の生成過程で見られるクラスター結合距離による保存，拡大，縮小といった空間のひずみに注目し，これらのひずみが階層的分類の各種手法に現れるための一般的条件を調べると共に，手法相互の関係を数理的に明らかにしている．この結果に基づき，新たな階層的分類法（一般化可変法）を提案し，これが既存のいくつかの手法の一般化に相当することなどが示されている．

第Ⅲ部（６章～８章）では，多変量混合分布モデルを分類法として用いる際の主な留意点（ＥＭ法の初期値設定とコンポーネント数の推定）について，実用的な分類方式を提案している．

ここで提案する分類方式は，第Ⅱ部で考察した階層的分類法および k-means 法などの，従来利用されてきたクラスター化法を，正規混合分布モデルの初期値設定の分類法として利用する．これは，データに内在する構造の特徴に応じて，各々の手法が固有の分類結果を与える（クラスター化の過程が異なる）という性質を利用して，初期値設定のための様々なクラスター化の状況を作り出すことに相当する．そして，これらの分類結果を多変量混合分布モデルの初期値設定に適用する．提案する分類方式は，事後確率（各コンポーネント分布への所属確率），判別率等の指標を用いることにより，得られた分類結果の客観的な比較を可能とし，結果として分類対象のデータ構造のより具体的な診断の手がかりが得られるという利点がある．従来行われてきた多変量混合分布の分類法は，初期値設定が分類結果におよぼす影響の評価は困難であるとされてきたが，ここに提案する方法は，これらの弱点を補うものである．

さらに，情報量規準を用いて正規混合分布モデルのコンポーネント数の推定を行う手続きを提案する．これは提案した分類方式と，ブートストラップ法でバイアス推定を行う手続きを併せて行うところに特徴がある．また，ブートストラップ標本から混合分布モデルのパラメータ推定を行う際の初期値の設定方法について考察し，この方法の有効性を数値実験により検証する．そして，線形近似によるブートストラップ・バイアス推定の変動減少法が，コンポーネント数の推定に際してブートストラップ反復回数の減少を可能とし，併せてＥＭ法の収束の遅さを補う方法としても有効であることを数値実験から観察する．

さらに多変量正規分布より裾の重いデータへの対処方法として，正規混合分布モデルの自然な拡張としての t混合分布モデルを提案する．ここで，t混合分布モデルは正規混合分布モデルを包含し，より一般的なモデルとして表現が可能になるという利点を示した．

第Ⅳ部（９章，10章）では，提案した分類方式の有用性を検証するために，いくつかの

事例データ解析を行う．

第一の例は，昆虫学の形質に基づく分類法との対比で，興味ある知見が得られた事例である．扱うデータセットはオーストラリアにおける野外調査で計測された「キバハリアリ」の計量データおよび形質データである．これを解析した研究者らは，主に形質データを利用した分岐分類を行い，所与のデータセットが９つの種群からなると結論づけた．

ここでは提案した分類方式を計量データに適用し，種群に相当するコンポーネント分布のその数の推定や，各々の個体（アリ）のコンポーネント分布への所属確率などを求める．この結果を上述の９種群と比較し，興味ある知見が得られた．とくに解析に用いる変数（特性）の選択，データの加工（比率変換など）を含めて，ここで提案した分類方式は，形質に基づく伝統的な分岐分類や系統解析等に先立つ事前処理法として利用できる．この意味で分類結果得られた群（クラスター）の客観的な情報は有用であるとの専門家の意見を得ている．

次に，LANDSAT の画像データへの適用例を取り上げる．ここで提案した分類方式のコンポーネント分布を多変量 t分布として分類を行う．これを正規混合分布モデルとの比較検討することにより，t混合分布モデルの有効性を示す．次に，各画素上の多変量特性データの事後確率，確率密度などの情報を用いて，分類結果を効果的に色彩画像化するための配色アルゴリズムを提案する．これら一連の手続きの特徴として，(1) 推定したコンポーネント分布の重なりあう様相が色彩画像として視覚化される，(2) コンポーネント分布やそれらの間の構造が画像上で色彩イメージとして視覚化され，分類結果の画像の解釈が容易になる，(3) 扱う画像データの画素数が比較的大きく（数十万～数百万程度），教師データ（トレーニング・データ）となる地上の詳細な情報が入手困難な場合などに有効である，などが挙げることができる．

論文の審査結果の要旨

審査委員会（駒澤勉，柳井晴夫（大学入試センター），小西貞則（九州大学大学院数理学研究科），村上征勝，馬場康維，大隅昇）は数物科学研究科統計科学専攻の中村永友君の論文について、数物科学研究科における課程博士の授与に係る論文審査等の手続き等に関する規程に基づき、公開の論文発表会（平成７年１月27日）を開催し、出席者・専門家等からの評価・質議を踏まえて本委員会は下記の論文の評価と試験結果により博士論文として十分な内容を備えていると判断した。この結果を数物科学研究科教授会（平成７年２月22日）の課程博士合否判定審査に報告し、課程博士（学術）として合格の審判判定を得た。

１．論文の評価

本論文の特色は，新たな分類方式を従来のアプローチとは異なる観点から構築し，これを実際データの分類方式に適用して，実用的にも有用であることを実証的に示したことにある．まず，探索的データ解析手法として広く利用されている組み合わせ的階層分類法に付随する空間のひずみを数理的に明らかにしたことは特筆すべきことである．従来のこの種の研究は発見的かつ主観的考察に終始していたが，これを数理的見地から体系的に考察し，しかも新たな階層的分類手法を誘導したことは高く評価される．つぎに，多変量混合分布モデルの抱える諸問題のうち，パラメータ推定のための初期値設定方法，分類結果の統計的評価方式，コンポーネント数推定等の問題について理論的かつ数値実験的に考察し，また実用的見地から具体的な分類方式を構築したことはきわめて意義がある．まず，パラメータ推定のための初期値設定の改善案として，複数の階層的分類法を利用することに着目し，これを分類結果を評価するための指標（判別率）と巧みに結びつけることで，初期値設定が分類結果に及ぼす影響を統計的に評価する方法を提案したことは優れた成果である．さらに，情報量規準に基づく混合分布のコンポーネント数の推定方式を，初期値設定の改良法，ブートストラップ・バイアス推定の変動減少法などを取り入れることによって具体化し，これを実験的に検証したことは，従来のこの種の分類法ではみられなかった特徴である．また，これらの考察と数値実験を通じて，データ解析の現場では多変量正規混合分布よりも裾の重い多変量特性データが見られることに着眼し，こうしたデータの解析に対処する一つの方法として，多変量 t分布の混合分布モデルの推定方式を導き，これを実際データに応用する新たな分類方式を提案したが，これは今までにない顕著な成果として評価される．以上の成果が，実際データの解析において有効に機能するかを具体的に検証するために，いくつかの事例データ解析を行っている．データ解析の初動探査も含めて，きめ細かい周到なデータ解析手順が示され，提案の分類方式の適用場面が，どのようなデータ解析においてどのように有用であるかを具体的に示したことは，本論文で提案の分類手法の実用化に大きく寄与するものである．とくに LANDSAT 画像データの解析において，提案分類方式による分類結果を，コンピュータ・ディスプレイ上に色彩分類画像として表示するための独自の配色方式を提案し，多変量混合分布モデルに基づく提案分類方式がこうした利用面でも有効であることを例証したことも高く評価される．

2．試験結果

数物科学研究科における課程博士の授与に係る論文審査等の手続きに関する規程第９条に基づき試験を実施した．論文ならびに口頭発表の他に，統計科学の基礎知識，今後の研究に関する展望等についても口頭試問がなされた．また，研究成果は，和文学術誌（既に掲載済み，投稿中），英文学術誌（投稿中），リサーチメモランダム（英文），共同研究報告リポート等で報告されており，語学力（英語）についても十分と判断された．

この結果，出願者は統計科学および関連する分野に関し学位を授与するに十分な学識を有するものと判断し，合格と判定した．（試験実施　平成７年１月27日）