第5章　結論

5.1　本研究のまとめ

本研究では，教師なし学習と呼ばれるクラスター分析手法にファジィの概念を導入したファジィクラスタリングの手法において，不確定性を含むデータを扱うためのアルゴリズムの開発を行った．従来のファジィクラスタリングでは，不確定性を含まないデータに対して分類を行った結果，あいまいさを含む分類が得られた．同様の発想として，与えられるデータ自身にあいまいさが含まれる場合も考えられてしかるべきである．そこで，不確定性 (あいまいさ) を含むデータをファジィクラスタリングする手法を開発した．ただし，従来のクラスタリング手法では, データに不確定性を仮定しないのに対し，不確定性を含むデータを扱うためには，クラスタリング手法はより複雑なものにならざるをえない．実際，本論文で開発した不確定性を含むデータに対するクラスタリング手法は，従来の不確定性を含まないデータに対するクラスタリング手法よりも複雑な手法となっている．もし，より一般的に様々な不確定性を扱えるような手法を考えるならば，クラスタリング手法はより複雑になり，実際どのような手続きが行われているか手順を追ってみることは困難になっていく．つまり，より一般的なところで議論しようとすると抽象性が増した結果，概念としてはシンプルであるとしても，その手続き (実際のアルゴリズム) はより複雑となり，アルゴリズムの実現が困難，あるいは計算時間が大きくなるなど，実際に応用する場面で利用されにくいものとなってしまう．

そこで，ある程度シンプルな部分で議論を行って，実際のクラスタリング手法の実現が容易で，そのアルゴリズムが行っていることがわかりやすいようなクラスタリング手法の開発を考えた．シンプルにする部分として，データは区間あるいは，三角ファジィ数の直積で表現されるとした．このようなデータでは，データの各々の成分ごとに距離を考えればよく，距離計算もより簡単で，見通しもよくなる．

また，クラスタリングアルゴリズムを開発するにあたってその根底においたのは，

ここで扱うファジィc-平均法は，評価関数を最適化することによって分類が行われる，ということである．不確定性を含まないデータに対するファジィc-平均法は，目的関数に2つの変数が含まれているために交互最適化という手法を用いて目的関数の最適化を行っている．そこで，本手法でもこの交互最適化を厳密に行うという視点で，アルゴリズムを開発した．

この結果，不確定性を含むデータに対するファジィクラスタリングアルゴリズムを実現することができた．

第2章では階層クラスタリングと非階層クラスタリングを概観した．クラスタリング手法のなかでもc-平均法は最もよく用いられ，この手法にファジィな分類を与えるファジィc-平均法が提案されている．ファジィc-平均法には，標準的な手法とエントロピー正則化を用いた手法がある．ファジィクラスタリングでは，ユークリッド距離の二乗が多く用いられているが，L_1距離も用いられている．区間あるいはファジィ数の直積からなる不確定性を含むデータを扱う場合，L_1距離はより自然な距離のとり方である．目的関数に含まれる不確定性を含むデータに対する距離を最長距離と最短距離を用いて定める．この2種類以外の距離としてハウスドルフ距離が考えられるが，本論文では，クラスター中心は不確定性を含まないと仮定するので，最長距離と一致することが示された．

第3章では，区間データに対するファジィクラスタリングのアルゴリズムを開発した．区間データに対するファジィc-平均法の目的関数を定めるために，最長距離と最短距離を用いて，区間データとクラスター中心の距離を定義した．目的関数をファジィc-平均法のアルゴリズムで最適化するために，クラスター中心の最適解の解法が必要となる．不確定性を含むデータに対する目的関数は，クラスター中心に関して凸であることから，偏導関数が負から正に転じる場所を発見するアルゴリズムを開発した．その計算量は，アルゴリズムを始める前のソーティングを除けば，従来のファジィc-平均法とオーダーが同じである．

第4章では，三角ファジィ数の直積で表されるファジィデータを扱った．このファジィデータのαカットは区間の直積なので，αカットとクラスター中心との間の距離は，3章と同様に最長距離あるいは最短距離を用いて定まる．この距離をαに関して積分することにより，ファジィデータとクラスター中心の間の距離が定義される．メンバシップとクラスター中心の2つの変数が含まる目的関数をファジィc-平均法のアルゴリズムで最適化するが，クラスター中心に関する最適解を求めるため

に L_1 距離に基づくアルゴリズムを開発した．このアルゴリズムは，区分線形なクラスター中心に関する偏導関数を探索するアルゴリズムである．アルゴリズムの計算量は，従来のファジィ c-平均法とオーダーが同じであり，計算量の面で劣るわけではない．ユークリッド距離の二乗に基づく場合は，非線形な偏導関数の探索が必要で，複雑であるため扱わなかった．

5.2　展望と今後の課題

実データに基づく手法の検証

開発した手法では，不確定性を含むデータに対して最長距離を用いた場合，最短距離を用いた場合の2種類の結果が得られる．また，各々の成分に含まれる不確定性を何らかの値で代表することにより，不確定性を含まないデータとして扱った場合の従来のクラスタリング手法を行うと，さらにもう1つの結果が得られる．これら3つの結果を比較することができる．すでに分類がわかっている不確定性を含むデータが与えられた場合には，クラスタリングを行った結果の，誤分類の数を従来の手法と本手法で比較することができる．また，データの傾向が知られているような不確定性を含むデータに対して，従来の手法と本手法におけるクラスタリング結果が得られた場合に，これらの結果におけるデータの傾向を比較することができる．このような比較により本手法の性質が明らかになっていく．このためには，実データを用いた多くの数値実験が必要となる．

他の手法への応用

ファジィ c-平均法では，メンバシップに対する制約として一般的に確率制約が用いられ，本稿でもこの制約を用いた．これ以外にファジィ c-平均法の制約条件を可能性制約に変えた可能性クラスタリング [38] において，不確定性を含むデータを扱うことが考えられる．このとき，ここで扱った区間データとクラスター中心あるいは，ファジィデータとクラスター中心の距離の定義はそのまま利用できる．また，非階層クラスタリングだけでなく階層クラスタリングにおいて区間データやファジィデータを扱う手法を考えることができる．

クラスター中心に対する不確定性の仮定

クラスター中心は不確定性を含まないと仮定して議論を行ってきたが，仮にクラスター中心に区間あるいはファジィ数の不確定性を許容したとしても，最適化の結果として不確定性を含むクラスター中心は得られない．ファジィc-平均法では，クラスター中心とデータとの間の距離を小さくすることにより，目的関数は減少する．クラスター中心に不確定性を仮定して，クラスター中心と個体との間の距離を最小にする最適化を行うと，最長距離を用いた場合は，不確定性を含まないクラスター中心が最適解となり，最短距離を用いた場合は，クラスター中心の不確定性を無限大にすることで，目的関数は0になるので，最適解が得られない．つまり，本論文の枠組みで，最長距離や最短距離を用いる限りにおいて，クラスター中心に不確定性を仮定する意味はない．ただし，最長距離あるいは，最短距離以外の距離を用いることや，ファジィc-平均法以外の別の目的関数を用いることにより，ファジィなデータをファジィクラスタリングした結果，ファジィなメンバシップとファジィなクラスター中心が得られるといった手法があるかもしれない．

不確定性を含むデータに対するデータ解析は，あまり研究されていない領域で，研究の余地が多くある．ここで述べた手法以外にも様々な手法やアルゴリズムが考えられる．また，多様な結果が得られることで，結果の解釈についても様々な可能性がある．このように，理論的にも実用的にも発展の余地が大きく，今後のさらなる研究が望まれる．