

**科学研究費助成事業 研究成果報告書**

平成 27 年 5 月 25 日現在

機関番号：32619

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500282

研究課題名(和文) 精度保証付数値計算技法を用いたクラスタ解析の初期値依存性と不確実性の包括的取扱

研究課題名(英文) Solving local convergence problem and missing value problem in clustering by validated numerics

研究代表者

神澤 雄智 (Kanzawa, Yuchi)

芝浦工業大学・工学部・准教授

研究者番号：00298176

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究の目的は、精度保証付き数値計算とその周辺の技法をクラスタ解析に導入し、主に初期値依存性問題を解決することと、データの不確実性を包括的に扱うことである。

初期値依存性問題について、精度保証付き数値計算における全解探索アルゴリズムはエントロピー正則化ファジィc-平均法(eFCM)に対して実時間で解を得ることができなかった。これはeFCMが多くの局所解を持つためである。一方で、Bezdek型ファジィc-平均法の最大化モデルからトレース最大化問題が自然に導かれることが明らかになった。データの不確実性については、欠損値を無限幅の区間で表して全解探索を援用するアルゴリズムを開発した。

研究成果の概要(英文)：The purpose of this research project is to introduce validated numerics techniques to clustering, to solve the local convergence problem in clustering and to deal with the uncertainty of given data for clustering.

In local convergence problem, the all solution algorithm in validated numerics could not produce its result for entropy-regularized fuzzy c-means (eFCM) within adequate time because eFCM has many local optimal solutions. On the other hand, it was clarified that a maximizing model of Bezdek-type fuzzified c-means algorithm can be transformed into a trace maximizing problem naturally, which can be solved globally.

In dealing with the uncertainty of give data, an algorithm was constructed to represent given missing values as the interval with infinite width, and to execute clustering along with divide such the intervals with the aid of all solution algorithm in validated numerics.

研究分野：クラスタリング

キーワード：ファジィクラスタリング

### 1. 研究開始当初の背景

クラスタ解析またはクラスタリング(与えられたデータを幾つかのグループに分類する手法の一つ)に関する研究は、学問上の興味対象となってきただけでなく、パターン認識をはじめとするさまざまな場面で活用されているが、まだ解決されていない問題点も多い。

クラスタ解析において K 平均法などの多くの手法は非凸非線形最適化問題に基づいているために初期値依存性を持ち、必ずしも大域的最適解が得られないので、通常はモンテカルロ法のように何種類もの初期値に対してアルゴリズムを試行して得られた解候補群の中から目的関数値が最も小さいものを選ぶ(マルチスタート戦略)。良い(と思われる)初期値を選ぶ方法が幾つか提案されているが、小手先の応急処置に過ぎない。また、スペクトラルクラスタリングなどの、目的関数の凸緩和問題を解く方法が提案されているが、得られた緩和解からクラスタ結果を解釈するのが難しく、結果として問題を先送りしているだけである。

クラスタリングにおいてデータが物理量の測定から得られたり人間の主観を数値化したものであれば当然、不確実性を考慮する必要がある。先行研究の多くは不確実性を表す区間などの集合としてデータを表したものの、アルゴリズムの途中で結局は決定論的議論に帰着させてしまっていて不確実性を論じ切れていない。また、集合内から標本を多数回サンプリングしてクラスタリングアルゴリズムを実行する研究も行われているが、データのバラツキを検討するには非常に莫大なサンプリング数を要するし、有限回サンプリングしたところで不確実データに潜む無限性を保証したことはない。

### 2. 研究の目的

本研究の目的は、精度保証付き数値計算とその周辺の技法をクラスタ解析に導入し、主に初期値依存性問題を解決することと、データの不確実性を包括的に扱うことである。

初期値依存性問題については、精度保証付き数値計算技術に基づく厳密大域最適化アルゴリズムを開発し、その性能を明らかにすることを目的とすると共に、クラスタリング手法が基とする最適化問題から適切な初期値設定指針を与える手法を構築することを目的とする。

不確実性の扱いについては、データの不確実性を表す領域や存在可能性などを、従来法のようにアルゴリズム途中で縮約することなく、不確実性を保ったままクラスタリングアルゴリズムを遂行する手法を構築することを目的とし、さらに、データの不確実性が互いに創刊を持つ場合を扱うために、アフィン形式などの高度集合表現を用いて不確実性を表し、高度集合表現データをクラスタリングするための手法を構築することを目的とする。

### 3. 研究の方法

初期値依存性問題については、精度保証付き数値計算技術に基づく厳密大域最適化アルゴリズムを開発し、その性能を明らかにする。本研究代表者が開発してきた非線形方程式の全解探索アルゴリズムやそれに関連する精度保証付き数値計算技術を非凸非線形最適化問題の厳密大域最適化アルゴリズムとして、人工データや実データに適用し、計算量評価と共に、局所解の特性を実験的に評価していく。また、クラスタリング手法が基とする最適化問題から適切な初期値設定指針を与える手法を構築する。クラスタリング手法が基とする最適化問題にできるだけ自然な最小の仮定を加えてトレース最大化問題のような唯一解を得られる問題に等価変換することを検討する。

不確実性の扱いについては、データの不確実性を表す領域や存在可能性などを、従来法のようにアルゴリズム途中で縮約することなく、不確実性を保ったままクラスタリングアルゴリズムを遂行する手法を構築する。これについても本研究代表者が開発してきた非線形方程式の全解探索アルゴリズムやそれに関連する精度保証付き数値計算技術を援用してデータの不確実性を最後まで保持するアルゴリズムを構築する。得られたアルゴリズムを実装し、人工データや実データなどを用いて計算量やクラスタリング精度を評価する。さらに、データの不確実性が互いに創刊を持つ場合を扱うために、アフィン形式などの高度集合表現を用いて不確実性を表し、高度集合表現データをクラスタリングするための手法を構築する。得られたアルゴリズムを計算機に実行し、人工データや実データなどを用いて計算量やクラスタリング精度を評価すると共に、区間を用いた場合との比較検討を行う。

### 4. 研究成果

初年度は、初期値依存性問題について、単純なベクトルデータに対して厳密大域最適化アルゴリズムを開発することから始めた。標準的ファジィ c-平均法に基づくアルゴリズムについては個体とクラスタ中心が一致した場合に特異な状況が発生し、これまでの近似計算では見過ごされてきた、厳密計算における問題点を明らかにした。エントロピー正則化ファジィ c-平均法に基づくアルゴリズムについてはその定式化を完了して一応の実装を済ませたが、単純なデータに対してでさえも収束に至らなかった。これは擬似的な極値や鞍点が膨大に存在するためと予想しているが論拠ある考察までには至らなかった。不確実性の包括的取り扱いについては、膨大なクラスタリングアルゴリズムの中でその対象となる手法を限定していく過程で、エントロピー正則化関係性ファジィ c-平均法が非ユークリッド関係性データにもそのまま適用可能であることと、関係性ファジィ c-平均法とカーネルファジィ c-平均法が実

際には同じ枠組みで議論できることを明中にするなど幾つかの成果を国際会議で発表した。これらの成果により、膨大なクラスタリングアルゴリズムを幾つかの枠組みに集約化することができ、不確実性を導入していく準備を整え始められた。さらに、単純なデータに対する標準的ファジィ c-平均法による結果だけであるが、区間演算を用いることによってクラスタ境界において帰属度がファジィであるために一概にクリस्प化できないことを明らかにした。

次年度は、初期値依存性問題について、可能性クラスタリングにおけるマウンテンクラスタリング的アプローチが全データを初期クラスタ中心更新式を反復すると幾つかのクラスタ中心に集約することを踏まえて、一つ目のクラスタ中心が得られた際に精度保証付き数値計算技法を用いて収束半径を保証する手法を開発したが、期待に比べて保証された収束半径が小さいために計算効率化には寄与する見込みが薄いことを明らかにした。その一方、もう一度原点に立ち返ってファジィ c-平均法の目的関数を見直した。エントロピー正則化ファジィ c-平均法において予言類似度を基にした最大化問題が提案されているのに対して、標準的ファジィ c-平均法において予言類似度を基にした最大化問題が提案されていないことに着目して、標準的ファジィ c-平均法の目的関数をクリस्प c-平均法の正則化という観点から改訂して、得られた最大化問題のカーネル化したものに対して2つの条件を仮定すると固有値問題に帰着されることから初期値問題を解決できる見込みがあることを明らかにした。データの不確実性の包括的取り扱いについては、線形計画法および二次計画法を用いた個体毎 スプレッドによる非ユークリッド関係性データクラスタリング、標準的ファジィ c-平均法を正則化の観点から見直すことによって得られる冪乗正則化ファジィ c-平均法と一般化標準的ファジィ c-平均法、三角不等式を基にした欠測関係性データの埋め込みによる共クラスタリング、カーネル解析に基づく関係性データの線形クラスタリングを開発した。

最終年度ではまず、前年度までに得られた次の成果を査読付き論文誌において発表した：(1)球面データに対する Bezdek 型ファジィ c-平均法の最大化モデルをカーネル化することによって、より一般的なデータに適用できるように拡張した。(2)(1)の成果をさらに推し進めて、初期値依存性を解決するためにトレース最大化問題に落とし込んで、その固有ベクトルを初期値に用いる手法を提案した。(3)冪乗正則化に基づく可能性クラスタリングを提案し、提案手法に対してカーネル密度推定に基づく初期値の設定指針とカーネル幅決定手法を提案した。不確実性に関する研究については、欠損値を無限幅の区間で表し、区間演算を用いた非線型方程式の全

解探索アルゴリズムに用いられる区間分割手法を適用しつつ、小区間に対してクラスタリングする手法を開発した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 8件)

Yuchi Kanzawa, Sequential Cluster Extraction Using Power-Regularized Possibilistic c-Means, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有、Vol.19, No.1, 2015, pp.67-73

<https://www.fujipress.jp/JACIII/>

Yuchi Kanzawa, On Kernelization for a Maximizing Model of Bezdek-like Spherical Fuzzy c-Means Clustering, Lecture Notes in Computer Science, 査読有、Vol.8825, 2014, pp.108-121

<http://link.springer.com/>

Yuchi Kanzawa, Relational Fuzzy c-Lines Clustering Derived from Kernelization of Fuzzy c-Lines, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有、Vol.18, No.2, 2014, pp.175-181

<https://www.fujipress.jp/JACIII/>

Yuchi Kanzawa, Fuzzy Co-Clustering Algorithms Based on Fuzzy Relational Clustering and TIBA Imputation, 査読有、Vol.18, No.2, 2014, pp.182-189

<https://www.fujipress.jp/JACIII/>

Yuchi Kanzawa, Generalization of Quadratic Regularized and Standard Fuzzy c-Means Clustering with respect to Regularization of Hard c-Means, Lecture Notes in Computer Science, 査読有、Vol.8234, 2013, pp.152-165

<http://link.springer.com/>

Yuchi Kanzawa, Relational Fuzzy c-Means and Kernel Fuzzy c-Means Using an Object-Wise beta-spread Transformation, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有、Vol.17, No.4, 2013, pp.511-519

<https://www.fujipress.jp/JACIII/>

Yuchi Kanzawa, Hard and Fuzzy c-Means Clustering with Conditionally Positive Definite Kernel, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有、Vol.16, No.7, 2012, pp.825-830

<https://www.fujipress.jp/JACIII/>

Yuchi Kanzawa, Entropy-Regularized Fuzzy Clustering for non-Euclidean Relational Data and for Indefinite Kernel Data, Journal of Advanced Computational Intelligence and Intelligent Informatics、査読有、Vol.16、No.7、2012、pp.784-792  
<https://www.fujipress.jp/JACII11/>

〔学会発表〕(計 10件)

Yuchi Kanzawa, On Bezdek-type Fuzzy Clustering for Categorical Multivariate Data, SCIS&ISIS2014、2014年12月4日、北九州国際会議場(福岡県・北九州市)

Yuchi Kanzawa, On A Maximizing model of Spherical Bezdek-Type Possibilistic c-Means and Fuzzy Multi-medoids Clustering, GrC2014、2014年10月23日、登別グランドホテル(北海道・登別)

Yuchi Kanzawa, A Maximizing Model of Bezdek-like Spherical Fuzzy c-Means Clustering, WCCI2014、2014年7月6日、北京(中国)

Yuchi Kanzawa, Relational Fuzzy c-Means and Kernel Fuzzy c-Means using a Quadratic Programming-based Object-wise beta-spread Transformation、2013年10月17日、KSE2013、ハノイ(ベトナム)

Yuchi Kanzawa, Comparison of Imputation Strategies in FNM-based and RFCM-based Fuzzy Co-Clustering、2012年11月24日、神戸国際会議場(兵庫県・神戸市)

Yuchi Kanzawa, FNM-based and RFCM-based Fuzzy Clustering for Tri-Relational Data、2012年11月24日、神戸国際会議場(兵庫県・神戸市)

Yuchi Kanzawa, Relational Fuzzy c-Lines Derived from Kernel Fuzzy c-Lines、2012年11月23日、神戸国際会議場(兵庫県・神戸市)

Yuchi Kanzawa, On Relational Fuzzy c-Means and Kernel Fuzzy c-Means Using Object-wise beta-spread Transformation、2012年11月21日、Girona(スペイン)

Kosuke Ihara, Yuchi Kanzawa, All Solution Algorithm for Parameter-Dependent Nonlinear Equations Using Affine Arithmetic、2012年10月22日、Palma(スペイン)

Yuchi Kanzawa, Yasunori Endo, On FNM-based and RFCM-based Fuzzy Co-Clustering Algorithms、プリズベン(オーストラリア)

〔図書〕(計 0件)

〔産業財産権〕  
出願状況(計 0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等  
なし

6. 研究組織

(1) 研究代表者

神澤 雄智 (KANZAWA, Yuchi)

芝浦工業大学・工学部通信工学科・准教授

研究者番号：090298176

(2) 研究分担者

なし( )

研究者番号：

(3) 連携研究者

なし( )

研究者番号：