

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 24 日現在

機関番号：12102

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500212

研究課題名（和文）推論則を用いた知識と最適化の融合によるクラスタ解析手法の高度化

研究課題名（英文）Development of Cluster Analysis Methods by Merging Knowledge and Optimization with Inference

研究代表者

遠藤 靖典（ENDO YASUNORI）

筑波大学・システム情報系・准教授

研究者番号：10267396

研究成果の概要（和文）：本研究では、まず、データやそれを扱う人間に含まれる不確実性を定式化した。次に、定式化した不確実性をデータ処理技法の基礎となる最適化問題に取り入れることにより、データ処理のモデルを構成した。その上で、構成したモデルに基づいて、特に不確実データに対する、クラスタ解析アルゴリズムを中心とするデータ解析のためのツールの高度化を行った。

研究成果の概要（英文）：We first formulated uncertainties included in data or human. Next, we constructed data process models by introducing the formulated uncertainties into optimization problems which are based on data processing through human inference. Third, we developed tools of data analysis, in particular, clustering algorithms for uncertain data.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	1,300,000 円	390,000 円	1,690,000 円
2010 年度	900,000 円	270,000 円	1,170,000 円
2011 年度	1,100,000 円	330,000 円	1,430,000 円
年度			
年度			
総計	3,300,000 円	990,000 円	4,290,000 円

研究分野：ソフトコンピューティング

科研費の分科・細目：情報学、感性情報学・ソフトコンピューティング

キーワード：ファジィ理論、クラスタリング

1. 研究開始当初の背景

データ解析、その中でも特に、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いくつかのグループに分類する手法の一つ）に関する研究は、古くから多くの研究者の学問上の興味の対象となってきただけでなく、パターン認識を中心とした分野への応用が期待されてきた。そして現在では、それらの分野のみならず、クラスタ解析を利用した yippy (<http://search.yippy.com/>) 等のデータ検索エンジン、医療現場におけるポジ

ترون CT による医療診断への応用など、当初想定されていたよりもはるかに広い分野において利用されている（例えば文献[1]）。しかし、かつてとは比較にならないほど多様化した価値観の中で、膨大な量のデータを処理しなければならない状況に置かれているにも関わらず、クラスタ解析に関する研究は、以前ほどの盛隆を見せていない。一つには、新しいクラスタ解析アルゴリズムを開発せずとも、現状で十分であるという理由もあろうし、また一つには、ある程度のアイデア

は出尽くし、新しいアルゴリズムを構築することに困難が伴っているという理由もあろう。しかし、現在我々が手にしているアルゴリズムでは、現在の複雑かつ大規模なデータの処理には限界があり、また、多くの人間の主観や、リスク・チャンス発見といった、さまざまな様相を持つ概念に対応することにも無理がある。特に、不確実性を持つような情報・データに関しては、これまでのクラスタ解析アルゴリズム（だけではなく、古典的な学問体系自体もそうであるが）では、実用的で適切な処理をすることはできないと本申請者は考えている。

不確実性に関する研究も古くから議論されてきたが、急速に脚光を浴びるようになってきたのはここ数年のことである。不確実性にはさまざまなフェーズがあるが、古典的な方法論だけではなかなか説明のつかない、人間に原因があるような確定的でない事象に対して、それらを解析・処理する、何らかの新しい方法論の開発が望まれており、これもまた、多くの研究者の興味を引くところとなっている。ファジィ・ニューロといったソフトコンピューティングに関する諸技法も、もともとはそれらの解析・処理を対象として生まれてきたといえる。これら人間の持つ知識を効果的に利用することにより、不確実性への効率的なアプローチが可能になるであろう。そして、人間の持つ知識を抽出・定式化することによって得られたデータベース（知識ベースと呼ぶ）と、データ処理技法の基本である最適化手法を、ファジィ推論に代表される柔軟な推論則を用いることにより融合させた、知識－推論－最適化によってデータ処理のモデルを構成し、構成されたモデルに基づいてアルゴリズムを組めば、特に不確実性を伴うデータ（不確実データ）に対して、これまでよりも柔軟な解析が可能になる（高度化）と思われる。

2. 研究の目的

そこで、本研究課題では、まず、データやそれを扱う人間に含まれる不確実性をモデル化し、人間の持つ知識を抽出・定式化することによって知識ベースを構築する。次に、ファジィ推論に代表される人間の持つ柔軟な判断を定式化した推論則を取り入れることにより、知識ベースをデータ処理技法の基礎となる最適化問題に融合させた、知識－推論－最適化によるデータ処理のモデルを構成する。その上で、構成したモデルに基づいて、特に不確実データに対する、クラスタ解析アルゴリズムを中心とするデータ解析のためのツールの高度化を行う。

まず、不確実性の概念を検討し直し、現実 に即した不確実性のモデルを構築する。次に、検討した不確実性のモデルをもとに、アンケ

ート・文献調査等を通じて、不確実性に対して人間が持っている知識を抽出し、さまざまなフェーズ毎に分類する。そして、それを知識ベースとしてデータベース化する。さらに、知識ベースと最適化手法を組み合わせるための推論則について、ファジィ推論をはじめとして検討する。その上で、推論則を用いてさまざまな最適化手法と知識ベースを組み合わせ、知識ベースと融合させるにはどのような推論則及び最適化手法が適しているか、また、どのように組み合わせるかを検討する。以上の検討をもとに、推論則を橋渡しとした知識ベースと最適化手法の融合を図り、それによるデータ解析手法、特にクラスタ解析アルゴリズムの高度化を行う。また、実際のデータを用いて、構築したアルゴリズムの検証を行う。

3. 研究の方法

以下に挙げる各研究内容について、番号順に実施した。

- (1) 本研究課題全体の詳細計画を確認した上で、協力を仰ぐ専門家、研究補助者への連絡を行い、本年度全体の計画が速やかに行われるための体制を確立した。
- (2) 不確実性のモデルを検討しなおした。不確実性のモデルにはさまざまな考え方があがるが、それらがすべて現実に即したモデルとは言えない。そこで、関連分野の専門家の指導を受け、現実に即したモデルを検討した。
- (3) データ解析において人間が利用する知識に関する調査を行った。
- (4) データ解析に適切な最適化手法の検討を行った。最適化手法には多くの方法が存在し、これまで非階層的クラスタ解析のベースとなってきた数理計画法にとどまらない。そこで、利用可能な最適化手法に関する文献調査を行い、どの最適化手法が不確実データを柔軟に扱えるかについて、数値計算を通じて検討した。
- (5) 検討したさまざまな最適化手法と知識ベースを組み合わせ、知識ベースと融合させるにはどのような最適化手法が適しているかを検討した。また、これらの手法と知識ベースとをどのように組み合わせるかについて議論した。
- (6) 以上の考察をもとに、知識－推論－最適化による不確実データに対するデータ解析手法、特にクラスタ解析アルゴリズムの開発を行った。特に、これまで提案されてきたクラスタ解析アルゴリズムを援用することによって構成できるかの検討を行った上で、全く新たなアルゴリズムを構成した。
- (7) 実際のデータを用いて、構築したアルゴ

リズムの検証を行った。

4. 研究成果

本研究における成果は以下の通りである。

- (1) データの不確実性の新たなモデル化を提案した。特に、ペナルティベクトル 2 次正則化という概念は、これまで提案されてきた許容範囲という概念より柔軟に扱うことができ、ペナルティベクトル 2 次正則化をクラスタリングの基本となる目的関数に組み込むことによって、許容範囲より容易に、データの持つ不確実性を最適化の枠組みで理論展開することが可能になる。
- (2) それらのモデル化を階層的クラスタリング、回帰分析等に導入し、新たなクラスタリングを構築した。また、ノンメトリックモデルに着目し、エントロピー正則化の観点から発展性の高いアルゴリズムを提案した。
- (3) 得られた成果を国内外の研究会・国際会議、ジャーナルへの投稿を通じて発表した。

最終的に本研究で開発したモデルの主要なツールは、ペナルティベクトル 2 次正則化という概念であり、この概念に基づいて新たに構築したアルゴリズムは、ファジィ c -平均法、階層的クラスタリング、回帰分析である。また、ノンメトリックモデルに関する新たな可能性を開いた。これらの新概念は、その有効性のみならず、数学的興味から、これまでの不確実性の扱いを大きく変えるものとして期待されるであろう。

ただし、不確実データに対するクラスタ解析アルゴリズムの高度化という大枠としてはその目的を達成したが、本研究の目的の 1 つである、推論則に基づくアルゴリズムの構築という観点からは必ずしも十分であったとは言いがたい。この点は今後の課題といえる。また、現在、ラフ集合に基づく不確実性を扱うクラスタ解析アルゴリズムの開発を行なっている。ラフ集合はブール集合より柔軟であり、ファジィ集合ほど過情報ではないため、不確実性を扱うツールとして過不足ないといえる。このラフ集合に基づくデータ解析アルゴリズムの発展が大きく期待される。

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 45 件)

- (1) Yasunori Endo : *On Entropy Based Fuzzy Non Metric Model -Proposal, Kernelization and Pairwise Constraints-*, Journal of Advanced

Computational Intelligence and Intelligent Informatics (JACIII), Vol.16, No.1, pp.169-173 (2012.1) (査読有), <http://www.fujipress.jp/finder/xslt.php?mode=present&inputfile=JACII001600010022.xml>.

- (2) Yasunori Endo, Yasushi Hasegawa, Yukihiro Hamasuna, Yuchi Kanzawa : *Fuzzy c -Means Clustering for Uncertain Data using Quadratic Penalty-Vector Regularization*, Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.15, No.1, pp.76-82 (2011) (査読有), <http://www.fujipress.jp/finder/xslt.php?mode=present&inputfile=JACII001500010009.xml>.
- (3) Yasunori Endo, Yukihiro Hamasuna : *Fuzzy c -means clustering with mutual relation constraints: construction of two types of algorithms*, Proc. of KES'11 Proceedings of the 15th international conference on Knowledge-based and intelligent information and engineering systems, Vol.1, pp.131-140 (2011) (査読有), <http://dl.acm.org/citation.cfm?id=2041289>.
- (4) Yasunori Endo, Arisa Taniguchi, Aoi Takahashi, Yukihiro Hamasuna : *On Hard c -Means using Quadratic Penalty-Vector Regularization for Uncertain Data*, Lecture Notes in Artificial Intelligence, Vol.6820/2011, pp.126-138 (2011) (査読有), <http://www.springerlink.com/content/5lut033p1n1524m3/>.
- (5) Yasunori Endo, Kouta Kurihara, Sadaaki Miyamoto, Yukihiro Hamasuna : *Hard and Fuzzy c -Regression Models for Data with Tolerance in Independent and Dependent Variables*, Proc. of The 2010 IEEE World Congress on Computational Intelligence (WCCI2010), -, #367 (2010) (査読有), http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5584335&abstractAccess=no&userType=inst.
- (6) Yasunori Endo, Yukihiro Hamasuna, Ayaka Tagaya : *On Hierarchical Clustering for Data with Tolerance*, The 6th International Conference on

Modeling Decisions for Artificial Intelligence (MDAI2009), -, (2009) (査 読 有), <http://www.mdai.cat/mdai2009/program.php>.

- (7) Yasunori Endo, Yukihiro Hamasuna, Makito Yamashiro, Sadaaki Miyamoto : *On Semi-Supervised fuzzy c-Means Clustering*, 2009 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2009), -, (2009. 8. 20-24) (査 読 有), <http://ieeexplore.ieee.org/Xplore/defdeny.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D5277177%26userType%3Dinst&denyReason=-134&arnumber=5277177&productsMatched=null&userType=inst>.

〔学会発表〕(計 26 件)

- (1) 谷口 亜里沙, 遠藤 靖典, 高橋 あおい : 不確実データに対するペナルティベクトル 2 次正則化ハード c-平均法, 第 27 回ファジィシステムシンポジウム (FSS2011) (福井県, 福井大学, 2011. 9. 12-14).
- (2) 辻 竜佳, 遠藤 靖典, 濱砂 幸裕, 栗原 宏太 : 許容範囲付きデータに対する主成分分析, 第 27 回ファジィシステムシンポジウム (FSS2011) (福井県, 福井大学, 2011. 9. 12-14).
- (3) Yasunori Endo: *Clustering for Uncertain Data and Its Applications to Risk*, The 1st International Workshop on Soft Computing and Disaster Control (SocDic2011) (Surabaya, Indonesia, 2011. 6. 20).
- (4) 濱砂 幸裕, 遠藤 靖典 : 対制約を用いたクラスタライズ許容範囲付きデータに対するファジィ c-平均法について, 第 26 回ファジィシステムシンポジウム (FSS2010) (広島県, 広島大学, 2010. 9. 13).
- (5) 遠藤 靖典, 宮本 定明, 山城 牧人, 濱砂 幸裕 : 許容範囲付きデータに対するハード c-回帰モデル, 第 25 回ファジィシステムシンポジウム (FSS2009) (茨城県, 筑波大学, 2009. 7. 15).
- (6) 山城 牧人, 遠藤 靖典, 濱砂 幸裕, 宮本 定明 : 半教師付きファジィ c-平均法に関する一考察, 第 25 回ファジィシステムシンポジウム (FSS2009) (茨城県, 筑波大学, 2009. 7. 15).
- (7) 栗原 宏太, 遠藤 靖典, 濱砂 幸裕, 宮本 定明 : 許容範囲付きデータに対するファジィ c-回帰モデル, 第 25 回ファ

ジィシステムシンポジウム (FSS2009) (茨城県, 筑波大学, 2009. 7. 15).

〔その他〕

ホームページ等

<http://endo.risk.tsukuba.ac.jp/~endo/>

6. 研究組織

(1) 研究代表者

遠藤 靖典 (ENDO YASUNORI)
筑波大学・システム情報系・准教授
研究者番号 : 10267396

(2) 研究分担者

なし

(3) 連携研究者

なし